GOES-R and JPSS Proving Ground Demonstration at the Hazardous Weather Testbed 2025 Spring Experiment Final Evaluation

Project Title: GOES-R and JPSS Proving Ground Demonstration at the 2025 Spring Experiment – Experimental Warning Program (EWP)

Organization: NOAA Hazardous Weather Testbed (HWT)

Evaluators: National Weather Service (NWS) Forecasters, Storm Prediction Center (SPC), National Severe Storms Laboratory (NSSL), University of Oklahoma (OU), Cooperative Institute for Severe and High-Impact Weather Research and Operations (CIWRO)

Duration of Evaluation: 5 May 2025 – 6 June 2025

Prepared By: Kevin Thiel (OU/CIWRO and NOAA/SPC)

Submitted Date: 19 September 2025

Table of Contents:

1. Executive Summary

This report summarizes the activities and results from the Geostationary Operational Environmental Satellite R-Series (GOES-R) and Joint Polar Satellite System (JPSS) Proving Ground demonstration at the 2025 Spring Experiment, which took place virtually at the National Oceanic and Atmospheric Administration (NOAA) Hazardous Weather Testbed (HWT) in Norman, Oklahoma from 5 May to 6 June 2025. This year featured 17 National Weather Service (NWS) forecasters participating in the EWP experiment, along with 18 product developers and visiting scientists. All NWS forecasters were from Weather Forecast Offices (WFOs) in four NWS regions. This group evaluated three experimental GOES-R products in the real-time simulated short-term forecasting, decision support service (DSS), and warning environment of the Experimental Warning Program (EWP). Additionally, they used cloud-based instances of the second-generation Advanced Weather Interactive Processing System (AWIPS-II) and web-based interfaces to interact with the experimental products. Organizations represented by the visiting scientists included four NOAA Cooperative Institutes, three NOAA line offices, and one external organization.

Forecaster feedback during the evaluation was collected through daily and weekly surveys, daily and weekly debriefs, blog posts, a warning and DSS reporting form, DSS graphics, and informal conversations during the testbed. Typical feedback included suggestions for improving algorithm performance, display techniques, training topics, and awareness of product applications or limitations. All three products evaluated in 2025 were advancements of previous product iterations from the 2024 GOES-R/JPSS Proving Ground (Thiel 2024), with new research-to-operations objectives and questions from these improvements. The three products were the GOES Radar Estimation via Machine Learning to Inform NWP (GREMLIN) model, the Optical flow Code for Tracking, Atmospheric motion vectors, and Nowcasting Experiments (OCTANE) suite, and the Probability of Severe (ProbSevere) LightningCast model. The Storm Prediction Center (SPC) and HWT Satellite Liaison Kevin Thiel (OU/CIWRO and NOAA/SPC) provided overall project management and subject matter expertise for the HWT Satellite Proving Ground efforts. Technical coordination and support for AWIPS-II was provided by Justin Monroe and Jonathan Madden (OU/CIWRO and NOAA/NSSL).

2. Introduction

GOES-R Proving Ground demonstrations in the HWT provide users with a glimpse into the capabilities, products, and algorithms available on the GOES-R satellite series. The education and training received by participants in the HWT fosters interest and engagement with new satellite data and promotes the continued operational readiness of GOES-R data and products for the NWS. The HWT provides a unique opportunity to enhance research-to-operations and operations-to-research (R2O2R) by enabling product developers and visiting scientists to interact directly with operational forecasters, and to observe the satellite-based algorithms being used alongside standard observational and forecast products in a simulated operational forecast and warning environment. This interaction helps developers understand how forecasters use these products and what improvements might increase product utility in NWS operations. Participant feedback in the HWT has been invaluable to the continued development and refinement of GOES-R algorithms since its inception in 2009. Furthermore, the EWP (Calhoun et al. 2021) facilitates the testing of satellite-based products in the AWIPS-II data processing and visualization system currently used at NWS WFOs.

All three weeks of the testbed (5-9 May, 19-23 May, and 2-6 June) were conducted virtually using Google Meet and Slack, hosting four to seven NWS forecasters each week. Feedback from forecasters and developers in previous experiments (Thiel 2023, 2024) provided sufficient motivation to continue the facilitation of both in-person and virtual demonstration weeks, as was initially planned. However, federal-employees and NWS forecasters were unable to travel during the evaluation period. This made it necessary for the 2025 experiment to revert to a fully virtual format. Before the testbed forecasters were provided user guides, PowerPoint presentations, and online learning modules for each of the products demonstrated. Additionally, the Satellite Liaison worked with each product developer group to formulate R2O questions to guide their evaluation in the experiment.

The Monday of each week began with introductions, an orientation session, and product summaries from developers, followed by familiarizing forecasters with their cloud-based AWIPS-II instances. Tuesday, Wednesday, and Thursday began with a discussion between the developers and forecasters of the previous day’s operations. After a brief forecast discussion each day by the Satellite Liaison, forecasters were placed into two to three groups localized to various NWS WFOs (hereafter simulated WFOs) across the United States to begin operations. The simulated WFOs were positioned to maximize the probability of severe thunderstorm activity each day, and to provide adequate evaluation of all demonstrated products. Mock-DSS events were created for a majority of simulated WFOs to investigate how the experimental products could also be utilized in communicating hazards to NWS partners. In five cases during the experiment, the Satellite Liaison asked forecasters at a simulated WFO to avoid looking at all available radar fields (WSR-88D and MRMS) to assess how the utility and applications of the experimental products changed in this scenario. At the end of each operations period, forecasters were given a daily survey regarding product performance and utility during the day’s events. Forecasters were encouraged to fill out an online form after submitting a convective warning, along with any forecaster who wished to provide DSS messaging for their assigned event. Additionally, forecasters could create DSS graphics using the experimental products for their simulated partners, further showcasing their ability as a communication tool.

Forecasters viewed GREMLIN, OCTANE, and ProbSevere LightningCast data in the cloud-based instances of AWIPS-II for the in-person and virtual demonstrations. Prior to the testbed, AWIPS-II procedures were built by the Satellite Liaison and product developers for each product, so forecasters could quickly access the products and leverage best display practices as described in their training. Within operations forecasters had several tasks, such as building procedures of their own to integrate experimental products with the ones they currently use, discussing their interpretation and experiences with the subject matter experts and developers, writing blog posts, and issuing warnings and DSS messages. Discussions between forecasters, developers, and visiting scientists involved questions from all groups concerning best display practices and applications, along with feedback from forecasters of what they were observing in real-time. Forecasters also had the opportunity to create blog posts which were published online to the HWT EWP Blog (https://inside.nssl.noaa.gov/ewp/) and the GOES-R HWT Satellite Proving Ground Blog (http://goesrhwt.blogspot.com/).

Testbed activities Monday through Thursday began at 1 pm CDT (18 Z) and ended at 6 pm CDT (23 Z). On Friday, an end of week survey was sent to the participants in the morning, followed by a two-hour final discussion with developers, observers, and SMEs to summarize the week's events and encapsulate key product themes from 1 to 3 pm CDT. More details from this schedule can be found in Figure 1.

Virtual Demonstration Schedule

Hour (CDT)	12:00	12:30	1:00	1:30	2:00	2:30	3:00	3:30	4:00	4:30	5:00	5:30
Monday			Orientation			Product Training/Operations						Daily Survey
Tuesday			Discussion/ Forecast		Operations							Daily Survey
Wednesday			Discussion/ Forecast		Operations							Daily Survey
Thursday			Discussion/ Forecast		Operations							Daily Survey
Friday	Weekly Survey		Weekly Debrief									

Figure 1: An approximate schedule of the virtual GOES-R Proving Ground HWT Experiment, outlining the major activities from each day.

The collective feedback from the 2025 GOES-R Proving Ground (daily surveys, weekly surveys, blog posts, and discussions) are summarized in this report. Each of products evaluated in the following subsections begin with a summary of the product and its intended applications, followed by science questions from each product developer group. Next, applications and feedback from the forecasters are summarized across all three weeks of the experiment. These are supported by forecaster questions from the surveys, forms, and blog posts throughout each section. Product recommendations are listed at the end of each section as 'recommended', 'strongly recommended', and 'highly recommended' in an ascending order of apparent significance from the forecasters.

3. Products Evaluated

3.1 GREMLIN

The GOES Radar Estimation via Machine Learning to Inform NWP (GREMLIN, Hilburn et al. 2021) model was demonstrated again in the 2025 Satellite Proving Ground HWT. GREMLIN is a convolutional neural network which uses spatial information from three ABI bands (Band 7: 3.9 µm, Band 9: 6.9µm, and Band 13: 10.3µm) and the Geostationary Lightning Mapper (GLM) Group Extent Density to produce an estimation of MRMS composite reflectivity across the ABI CONUS and Mesoscale scenes for daytime and nighttime scenarios. The model is trained against MRMS composite reflectivity east of 105°W during the warm season, and its reflectivity estimates are constrained to values between 0 and 60 dBZ. GREMLIN was intended to provide a spatially uniform parameter for initializing convection with numerical weather models; however, it is expected that forecasters may leverage GREMLIN reflectivity estimations in areas of little-to-no radar coverage. Additionally, convective features from satellite imagery associated with initiating convection may precede those from radar to increase lead time. In the 2024 experiment GREMLIN was only available for Full Disk scenes, and participants strongly recommended more frequent updates from the ABI CONUS and Mesoscale scenes instead of the Full Disk scene.

Within the 2025 Satellite Proving Ground, GREMLIN data were generated from the GOES-East and GOES-West ABI CONUS and Mesoscale scenes, updating every 5 minutes and one minute respectively, with a horizontal resolution of 2 km. GLM Group Extent Density accumulations matched the update frequency of the CONUS and Mesoscale scenes at 5 minutes and 1 minute respectively. Before the experiment forecasters were provided a one-page QuickGuide highlighting the inputs, applications, and limitations of GREMLIN. AWIPS-II procedures were created to help answer the R2O questions listed below, including four-panel displays with GREMLIN on the CONUS and Mesoscale scenes, GLM Flash Extent Density, MRMS composite reflectivity, and ABI bands 7, 9, and 13. With these procedures, forecasters could interrogate GREMLIN model inputs alongside its output and validation fields.

GREMLIN R2O Questions

- Does GREMLIN provide useful information to increase confidence for issuing warnings or communicating hazards with the public in areas lacking good quality ground-based radar coverage?
- Does GREMLIN on ABI Mesoscale and CONUS scenes provide timely information for initiating convection when compared to local radar?
- When ground-based radar networks experience outages, how useful is GREMLIN in providing a gap filler to forecasters above and beyond other sources of information?

Use of GREMLIN in the HWT

In the pre-testbed survey, forecasters were asked which of the three experimental products they were most excited to use, along with the product they were most uncertain to use. Amongst the ten responses, GREMLIN received the most votes for both questions. Forecasters noted the potential utility of satellite-derived radar estimates in areas with little-to-no WSR-88D coverage, while expressing uncertainty in applying synthetic radar data and how infrequently they currently use composite reflectivity in their home office. Throughout the testbed forecasters compared composite reflectivity estimates from GREMLIN with MRMS composite reflectivity and explored how well GREMLIN identified convective features and trends. Forecasters were told GREMLIN's maximum reflectivity estimate is 60 dBZ, and frequently noted thunderstorms with reflectivity values close this maximum as signal of intense convection. GREMLIN unexpectedly produced values exceeding 60 dBZ in at least three cases of mature supercells in highly unstable environments during the experiment. Forecasters interpreted this signal from GREMLIN as a high confidence in overall thunderstorm intensity and severity potential.

One of GREMLIN's expected applications was to quickly identify initiating convection. In the weekly surveys all but three of the 16 participants stated they felt 'Somewhat confident' or 'Very confident' using GREMLIN to identify initiating convection in areas with little-to-no radar coverage. When asked a similar question, but with respect to issuing warnings and communicating hazards, again all but three forecasters responded with 'Somewhat confident' or 'Very confident'. From the 63 warnings that forecasters reported in the warning/DSS form, 22 mentioned that information from GREMLIN influenced their warning process. Forecasters mentioned increasing composite reflectivity values from GREMLIN, which coincided with thunderstorm intensification signals from ABI infrared bands, GLM flash rates, local radar, and OCTANE.

> 'Interesting feature with Gremlin that would possibly suggest that there is a hook on the storm noted below to the right [Figure 2]. Hence we issued a SVR with TOR possible tag.'
> 8 May 2025, Blog Post: *Greer Severe Weather Blog*
> https://inside.nssl.noaa.gov/ewp/2025/05/08/greer-severe-weather-blog/

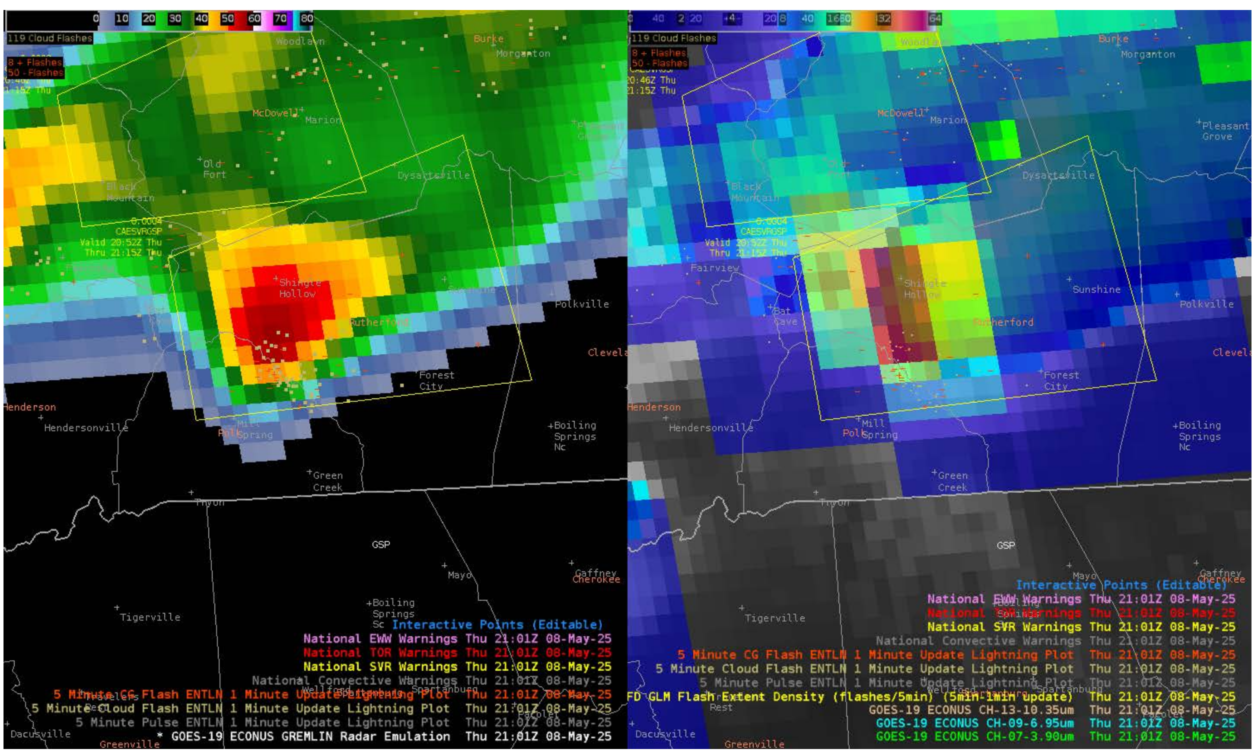

Figure 2: GREMLIN output from GOES-19 CONUS (left) when compared with GLM Flash Extent Density and ABI Ch. 13 Clean-IR Brightness Temperature from a thunderstorm in western North Carolina on 8 May 2025.

> ‘Without being able to use radar data, I thought GREMLIN did very well showcasing the cells compared to satellite. There were two specific cells that GREMLIN showed >60 dBZ in which we warned on’
> *NWS Forecaster – End of Day Survey*
>
> ‘the rhythm we fell into for warnings was using the GREMLIN product to track the storm cores and draw the polygon and cross-referencing with ground-based lightning networks to make sure parallax was accounted for.’ [Figure 3]
> 4 June 2025, Blog Post: *Tornado Warning St. Louis Without Radar*
> https://inside.nssl.noaa.gov/ewp/2025/06/04/tornado-warning-st-louis-without-radar/

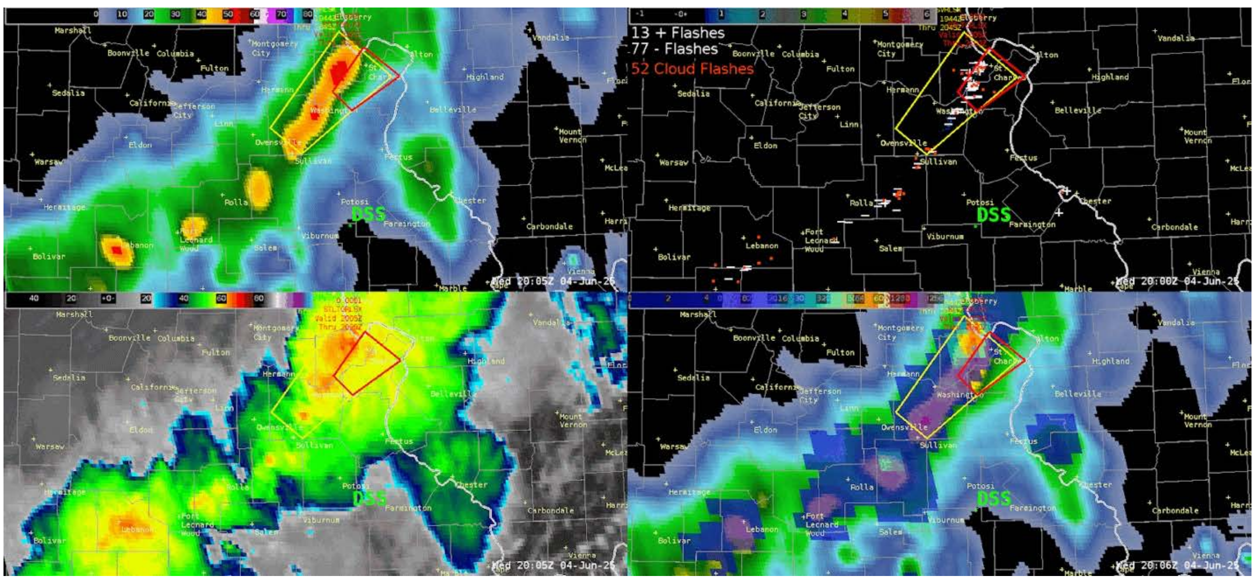

Figure 3: GREMLIN composite reflectivity from GOES-East MESO1 (upper left), NLDN CG flashes and ENTLN cloud flashes (upper right), ABI Clean-IR brightness temperature (lower left), and GLM Flash Extent Density overlaid on GREMLIN from GOES-East CONUS (lower right) near St. Louis, Missouri on 4 June 2025. Animation.

Use of GREMLIN reflectivity estimates were more frequent when forecasters were placed in CWAs with poor radar data and during radar denial cases. In the daily surveys participants were asked about the overall utility of GREMLIN. Of the five utility levels presented, approximately one-third (20/59) of forecasters found GREMLIN as ‘Moderately Useful’ (level three of five). Meanwhile 44% (26/59) responded with the lowest two categories of ‘Not at all Useful’ or ‘Slightly Useful’, and 22% (13/59) responded with the highest two categories of ‘Very Useful’ or ‘Extremely Useful’. When binning these responses against their perceived quality of radar data for the day (Figure 4), trends observed in operations were identified. Forecasters frequently stated the utility of GREMLIN in the lowest three bins when their perceived radar data quality in the greatest four bins from ‘Poor’ to ‘Excellent’. However, in the lowest radar quality bin of ‘Terrible/None’ all responses were focused the three greatest GREMLIN utility bins. These results were similar to those observed in the 2024 Satellite Proving Ground Experiment (Thiel 2024).

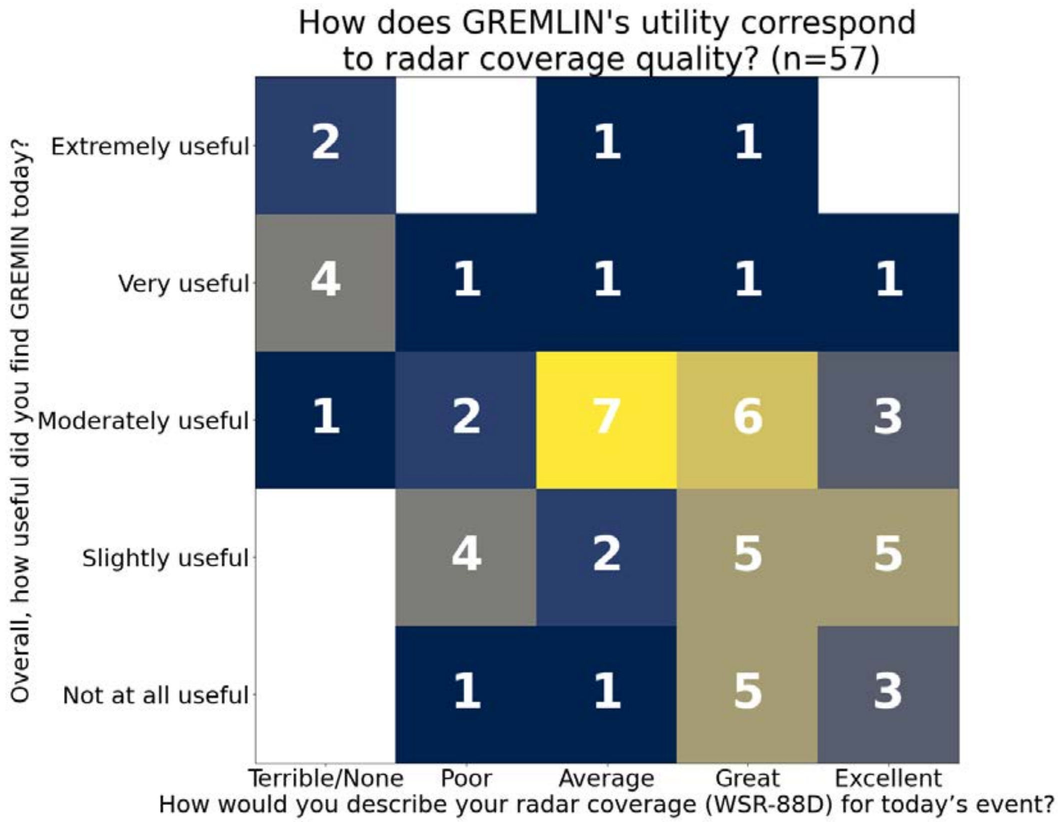

Figure 4: A two-dimensional histogram of daily survey responses (57) showing participant's apparent radar coverage quality and GREMLIN utility.

> 'I believe GREMLIN is a great supplemental product when used in tandem with satellite imagery and lightning data. It offers a very reasonable starting point for issuing warnings in areas of little to no radar coverage. It was also observed to provide a modest level of added confidence when nowcasting convective evolution, especially with higher emulated reflectivities were observed.'
> *NWS Forecaster – End of Week Survey*

> 'With no radar data available I feel GREMLIN can provide great value when used with other satellite products. However, there were some notable limitations like being too smooth (Not showing higher reflectivity like MRMS at times) and struggling to depict storms when other anvils or convective cloud debris interferes. If there is MRMS data available I will likely use that over gremlin DATA as it did better overall.'
> *NWS Forecaster – End of Week Survey*

For the first two weeks of the experiment, lightning information for the GREMLIN CONUS and Mesoscale scenes were accumulated every 5-minutes and 1-minute respectively. However, throughout both weeks forecasters noted in blog posts, group discussions, and surveys that the GREMLIN Mesoscale reflectivity values rapidly increased and decreased in areas with infrequent lightning activity. These jumps were most prominent in areas of stratiform precipitation. In the third week of the experiment, GREMLIN Mesoscale data was adjusted to accumulate lightning data every 5-minutes, and after this change no forecasters mentioned these issues. Additional conversations regarding the time integration of lightning data from longer, shorter, and multiple time integration windows was discussed as a future avenue of development.

Forecasters were asked in the daily surveys how well composite reflectivity estimates from GREMLIN compared to observed values from the MRMS data within convective cores. No forecasters found that GREMLIN values were greater than those from MRMS, while half (25/50) found that GREMLIN was 5 to 10 dBZ lower. The remaining responses were split between GREMLIN values that were about the same (+/- 5 dBZ) and much lower (<10 dBZ) compared to MRMS. Similarly, forecasters noted in blog posts, group discussions, and daily surveys instances where GREMLIN composite reflectivity values were reduced compared to MRMS composite reflectivity. Follow-up discussions included the impact of anvils from adjacent convection obscuring signals from GREMLIN, machine learning methods used to train the GRELMIN model, and the quality of GLM flash detections. Additionally, the lightning data used by GREMLIN data were improperly scaled during the first two weeks of the experiment, resulting in lower GLM group extent density rates for GREMLIN CONUS data by a factor of 0.5 and GREMLIN Mesoscale data by a factor of 0.1. This scaling reduced composite reflectivity estimates from GREMLIN and may have biased the forecasters when asked to compare between GREMLIN and MRMS composite reflectivity products in their daily surveys.

> ‘Overall, I thought the GREMLIN output looked too smooth. I was not really able to see the kind of sharp gradients in reflectivity or even smaller scale showers/storms that would signify a decision making point in warning operations. This is more of an observation than a suggestion since I know we're limited to the resolution of satellite data. But if there was any option to smooth less, I would take it.’
> *NWS Forecaster – End of Week Survey*

> ‘In the absence of radar data, GREMLIN appeared to provide a good alternative to traditional radar, especially when trying to draw warning polygons. The parallax issue created some challenges, but this was remedied some by lightning data.’
> *NWS Forecaster – End of Day Survey*

> ‘One feature that was very noticeable for our area compared to further south in the warm sector was the low-topped nature of our convection…GREMLIN returns were much weaker than one should have expected, even once lightning production got going in a cell… It seems like the low-topped convection and warmer cloud tops rather significantly hindered the performance of GREMLIN.’ [Figure 5]
> 20 May 2025, Blog Post: *GREMLIN Struggles with Low-Topped Convection*
> https://inside.nssl.noaa.gov/ewp/2025/05/20/gremlin-struggles-with-low-topped-convection/

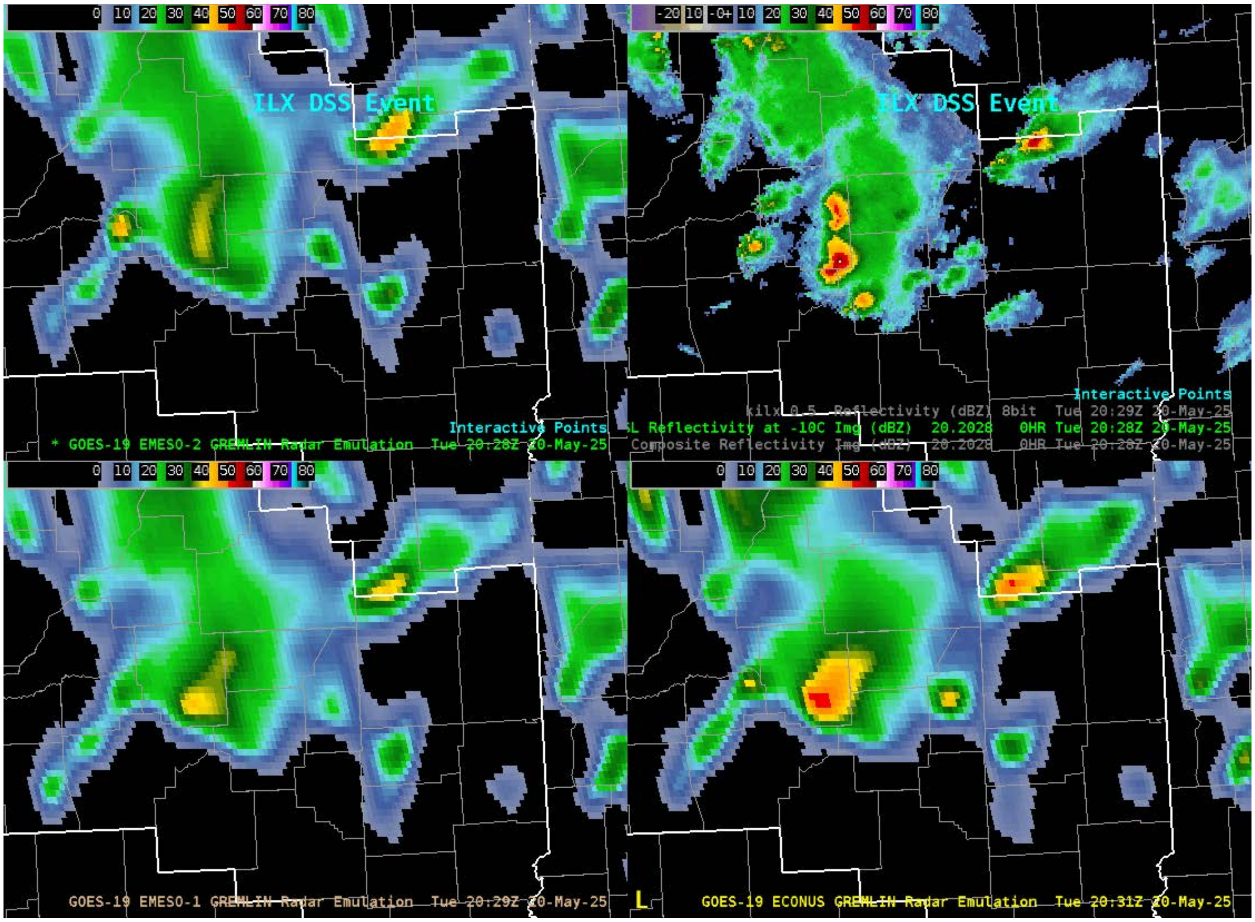

Figure 5: A comparison of GREMLIN Composite reflectivity from the GOES-East MESO-2 (upper left), GOES-East MESO-1 (lower left), and GOES-East CONUS (lower right) with MRMS composite reflectivity (upper right) in central Illinois on 20 May 2025. Animation.

In areas with little-to-no radar, forecasters tasked with issuing convective warnings expressed the desire for GREMLIN to supplement other reflectivity products besides composite reflectivity. Height-based reflectivity fields at constant altitudes or isotherms were suggested, along with vertically integrated reflectivity products such as vertically integrated ice (VII), vertically integrated liquid (VIL), and maximum expected size of hail (MESH). Forecasters issuing warnings also frequently requested a parallax corrected version of GREMLIN, and in group discussions they stated the importance of accurate polygons which reflect spatial extent of a thunderstorm's hazards while maintaining smaller false alarm areas.

> 'When using MRMS, I find vertically integrated ice (VII) and max estimated size of hail (MESH) to be very helpful. Could these be incorporated into GREMLIN?'
> *NWS Forecaster – End of Week Survey*

> 'Really the only issue I saw with it was the parallax problems. My CWA deals with a lot of flash flooding along terrain that radar doesn't necessarily reach and I see a product like this being extremely valuable someday.'
> *NWS Forecaster – End of Week Survey*

Recommendations for Operational Implementation

Based upon the evaluation of the GREMLIN model in the 2025 HWT Satellite Proving Ground, the following items have been recommended:

- **It is recommended that GREMLIN be used in areas with little-to-no radar coverage to identify initiating convection and monitor trends in thunderstorm intensity.** Other data sources such as satellite imagery, lightning data, and mesoanalysis fields should be used alongside GREMLIN to infer reflectivity estimates and increase confidence in observed trends.
- **It is strongly recommended that additional reflectivity products outside of Composite Reflectivity be produced, and a parallax correction be applied to the data for a more accurate spatial depiction of convective features and associated hazards.** MRMS products such as isothermal reflectivity, MESH, VII, and VIL may be considered
- **It is strongly recommended the data integration time of GLM group energy density data be explored to improve GREMLIN performance when identifying intense convection and thunderstorms with greater composite reflectivity values.**

3.2 OCTANE

The Optical flow Code for Tracking, Atmospheric motion vectors, and Nowcasting Experiments (OCTANE, Apke et al. 2022) suite of products were demonstrated again in the 2025 Satellite Proving Ground HWT. Near-pixel level brightness motions in cloud textures from the ABI visible band (0.64 µm) during the day and the clean-infrared band (13.3 µm) at night are calculated using consecutive images from the GOES-R series ABI. Magnitudes are calculated in m s^{-1}, a 5-minute median filter is applied to mitigate jitter signals, and new outputs are available with a latency of 2.5 to 5 minutes. OCTANE speeds and directions are calculated for the ABI Mesoscale scenes using the visible (infrared) band when the solar zenith angle is less than (greater than) 80 degrees with a resolution of 0.5 km (2 km) at nadir. 'Sandwich' imagery is created to leverage reflectance values from the ABI visible or infrared bands to display OCTANE speed and direction.

Cloud-Top Divergence (CTD) and Cloud-Top Cooling (CTC) fields are derived from time trends in OCTANE retrievals from the ABI Mesoscale Scene and ABI infrared imagery. Both products are also produced as a sandwich with visible imagery and attempt to provide more concise cloud-top information when compared to the OCTANE speed and direction fields. Based on feedback from the 2024 experiment, three display options of the OCTANE CTD product were provided with no smoothing, medium smoothing, and high smoothing to potentially remove excessive noise while keeping relevant convective signals. The suite of demonstrated OCTANE products provides wind information at cloud top within each mesoscale scene, highlighting environmental shear and cloud-top divergence (cooling) from developing and mature convection.

New for the 2025 experiment, OCTANE was run on ABI CONUS imagery to create synthetic MesoAnywhere imagery by interpolating ABI CONUS imagery every one minute from the OCTANE retrievals. The OCTANE MesoAnywhere product was available from the GOES-East ABI for bands 2, 5, and 13 (0.64µm, 1.6 µm, and 10.3 µm) and the Day Cloud Phase Distinction (DCP) RGB in a movable 3000 by 3000-pixel box (approximately 1500 km by 1500 km). For the 2025 experiment, all OCTANE products were made available in AWIPS-II through procedures created by the developers. The CIRA SLIDER webpage also provided a default view of the OCTANE speed sandwich along with the CTC and CTD sandwich.

OCTANE R2O Questions

- What scales of OCTANE cloud-top divergence were most useful for inferring deep convection intensity trends for convection warning operations?
- Did you find the Cloud-top Divergence product was more useful than the speed/direction sandwiches alone?
- Does the OCTANE MesoAnywhere product provide useful information for convection warning operations in the absence of GOES mesosector coverage?

Use of OCTANE in the HWT

Forecasters noted their interest in the OCTANE product suite in the pre-testbed survey, stating that the ability to quantify cloud-top winds and divergence trends was a new concept they felt could positively impact operations. Throughout the testbed's daily survey, forecasters stated that they used the OCTANE wind to identify updraft trends using divergence signals products for over 90% of shifts (52/55). Almost half used OCTANE winds to note initiating convection in sheared environments (27/55), and over one-quarter identified mesoscale ascent (16/55). Similar to the 2024 experiment, forecasters pointed out numerous examples of initiating convection and changes in updraft intensity through the speed, CTC, and CTD product signals. In cases where radar data were poor or unavailable, forecasters heavily relied on OCTANE speed and CTC alongside ProbSevere LightningCast to monitor for developing convection. They also used the speed, CTC, and CTD products alongside GREMLIN to monitor ongoing convection and updraft trends that may signal an intensifying or diminishing updraft. From the 63 reported warnings throughout the experiment, 45 (71%) referenced OCTANE as an influence in their warning decision.

> ‘the now mature primary supercell showed strong divergence within OCTANE and a wake near and immediately downstream of the overshooting top, indicative of a very strong and likely severe storm [Figure 6]…Overall, OCTANE aided our warning decisions significantly…the addition of satellite-derived storm top divergence was a good confidence nudger and helped us get half an hour to an hour lead time on large hail and over 15 minutes on a tornado. LightningCast and OCTANE helped guide us toward which cell would eventually become dominant during the initiation phase.’
> 5 June 2025, Blog Post: *New Mexico and Texas Tornado Outbreak*
> https://inside.nssl.noaa.gov/ewp/2025/06/05/new-mexico-and-texas-tornado-outbreak/

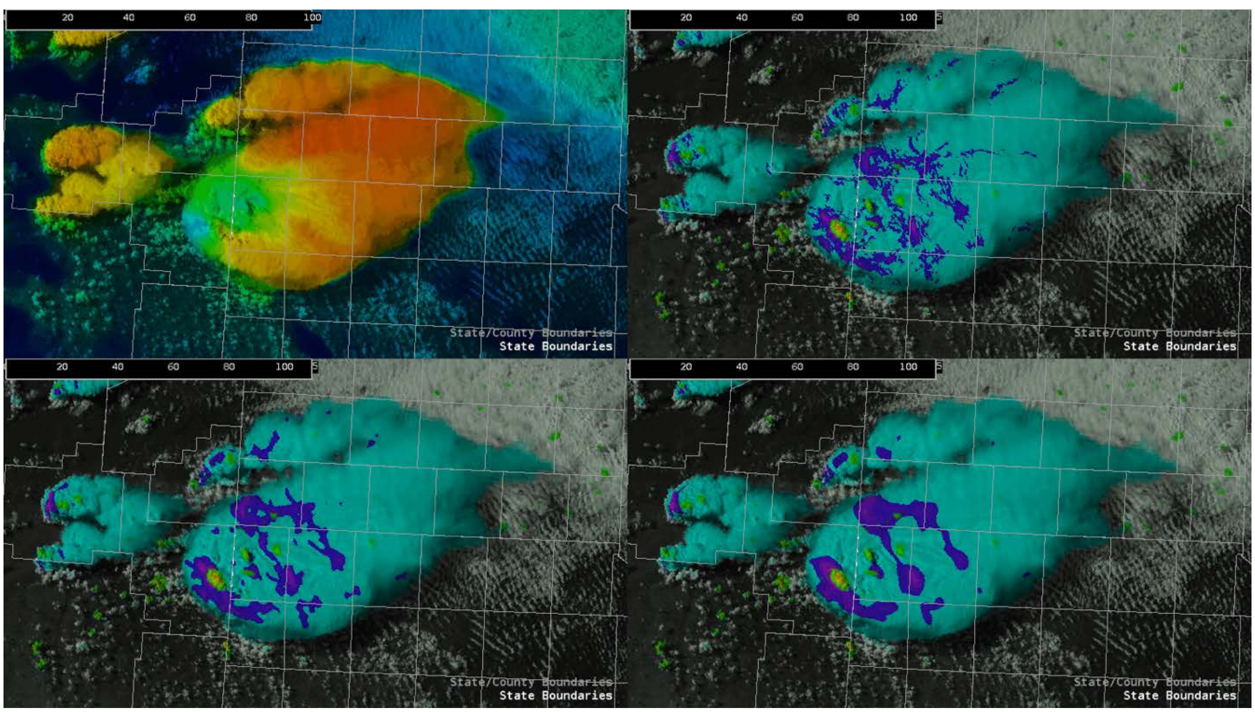

Figure 6: OCTANE Speed Sandwich (upper left) along with the OCTANE Cloud-Top Cooling and Divergence Sandwich with no smoothing (upper right), medium smoothing (lower left), and high smoothing (lower right) for a mature supercell on the Texas and New Mexico border on 5 June 2025. Animation.

> ‘OCTANE detected appreciable cloud-top cooling prior to a 1.25” hail report in Mississippi County, Arkansas. This cloud-top cooling indicated a strengthening updraft, which can be associated with an increasing hail threat in a favorable environment. The

most substantial cooling rate was approximately at 2005 UTC, which was 14 minutes prior to the hail report at 2019 UTC.' [Figure 7]
20 May 2025, Blog Post: *OCTANE Cloud-Top Cooling Prior to Hail Report*
https://inside.nssl.noaa.gov/ewp/2025/05/20/octane-cloud-top-cooling-prior-to-hail-report/

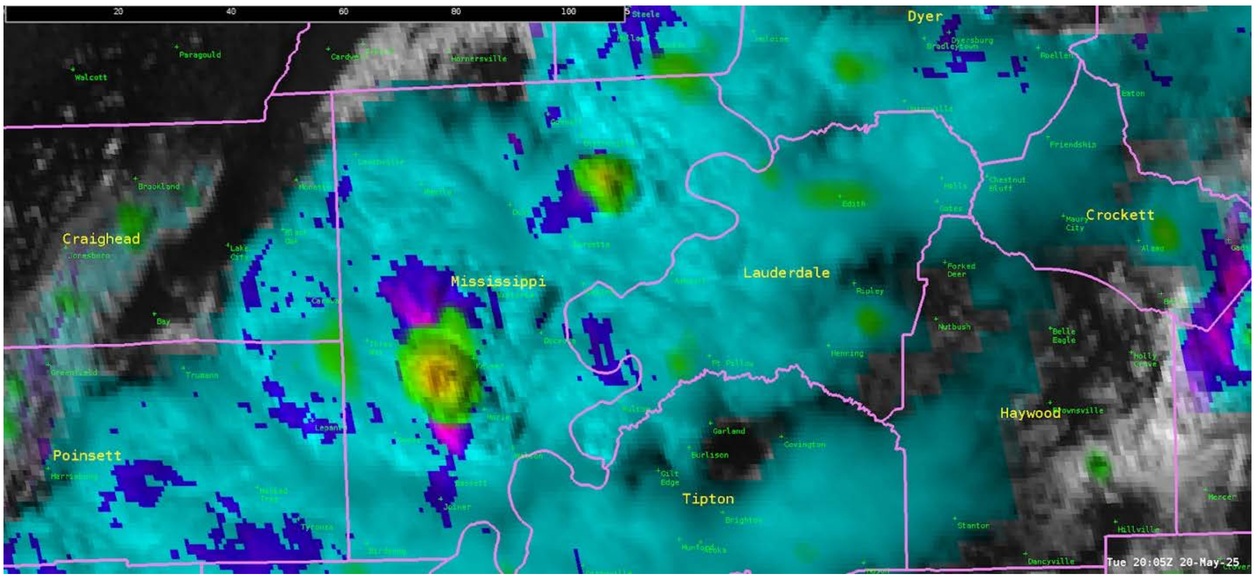

Figure 7: OCTANE Cloud-Top Cooling and Divergence Sandwich from a thunderstorm in Arkansas on 20 May 2025.

'OCTANE storm top divergence was extremely useful for gauging updraft intensity trends. We issued warnings that were heavily weighted using OCTANE divergence product.'
NWS Forecaster – End of Day Survey

'Used the Octane 4-panel with Cooling and Divergence. Especially liked the Octane Speed / visby that showed the low level stable clouds feeding into the storm. These stable clouds are tracers for streamwise vorticity right into the updraft. Radar was at about 12Kft, so not helpful with tornado warnings other than to say there was a supercell in a favorable environment for tornadoes.'
NWS Forecaster – Warning Report Form

Overall, forecasters preferred using the OCTANE CTD product to diagnose divergence rather than the OCTANE speed and direction sandwiches. Approximately 72% (31/43) of forecasters in the daily surveys found the divergence signals from the CTD product clearer than the speed and direction products, while around 23% (10/43) found them to be nearly identical. In group discussions, the participants frequently noted the OCTANE CTC and CTD products allowed them to quickly identify and diagnose convective trends and make proactive warning and DSS decisions. The following quotes from forecasters highlight key points made about the OCTANE divergence product when compared to the speed and direction products.

> ‘While one could visually tell strong divergence with ease using just the OCTANE speed sandwich, more moderate divergence was a bit tougher to assess with that product, but it was usually easy to see in the OCTANE divergence products. And having the Cloud-Top Cooling overlaid on the Divergence in certain procedures helped to easily see both fresh or renewed growth, in addition to the divergence.’
> *NWS Forecaster – End of Week Survey*

> ‘I found myself gravitating more towards the CTD and CTC panels. They were easier to comprehend, and were easier to quickly glance at to note the stronger storms’
> *NWS Forecaster – End of Week Survey*

Three levels of smoothing (None, Medium, and High) for the OCTANE Cloud-Top Divergence product were presented to participants through a four-panel procedure with all three levels present in AWIPS-II. Forecasters viewed the signals from each CTD option alongside the OCTANE Speed Sandwich during operations and compared how well each product signaled changes in updraft intensity. In the daily and weekly surveys forecasters found the CTD product with medium smoothing had the clearest divergence signal. This was followed by the ‘No smoothing’ option, and very few forecasters responded that they preferred the ‘High smoothing’ option. In group discussions quickly identifying features and trends at storm-scale, while not losing divergence details at storm scale, was emphasized for incorporating the OCTANE CTC and CTD products into their workflow when interrogating convection, especially when interrogating clusters of thunderstorms. A few forecasters did prefer the original ‘No smoothing’ option, which was more frequent later into the week once forecasters were more familiar with the CTD product and its interpretation.

> ‘The areas of CTD were few and far between enough that I wanted to see all of them with no smoothing. I didn't think it was noisy at all without smoothing and I didn't want to miss out on any small scale features that may have been smoothed over. It was also easy enough to edit the colors such that the higher end CTD values stuck out more and offered a better situational awareness.’
> *NWS Forecaster – End of Week Survey*

> ‘I do like the looks of the smoothing option, it is easier on the eyes and rids of any excess noise. I feel this helps to recognize trends easier and remain focused on those trends through the animation resulting in being particularly helpful in real-time operations…if given the choice, I’d lean slightly towards the Medium Smooth based on the High Smooth potentially smoothing out some important features and the No Smooth being too harsh’ [Figure 8]
> 6 May 2025, Blog Post: *OCTANE Cloud-Top Divergence Comparison*
> https://inside.nssl.noaa.gov/ewp/2025/05/06/octane-cloud-top-divergence-comparison/

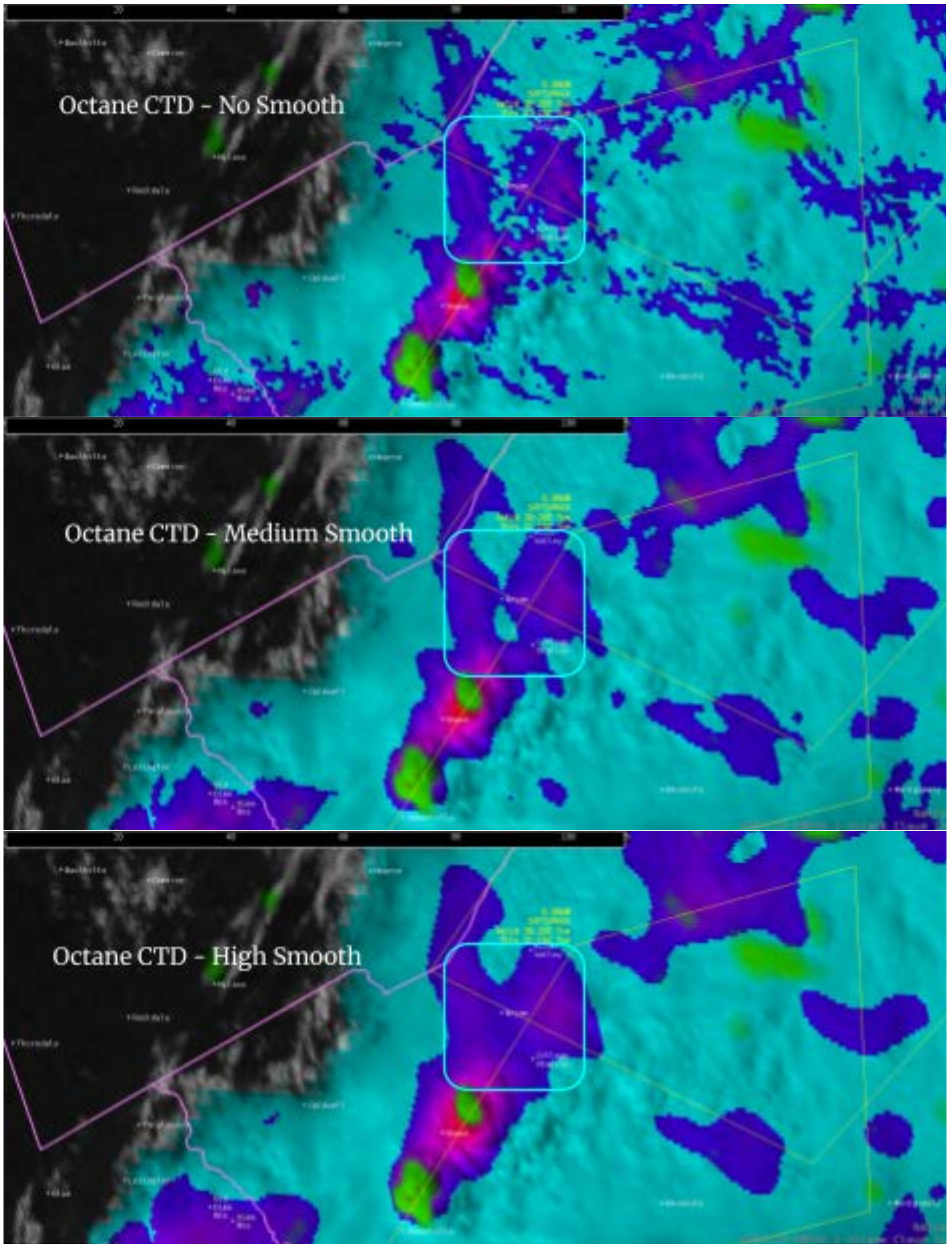

Figure 8: OCTANE Cloud-Top Divergence and Cooling Sandwich with no smoothing (top), medium smoothing (middle), and high smoothing (bottom) applied to the divergence field from 6 May 2025.

Forecasters in operations worked with OCTANE developers to understand, and in some cases modify, displays of the OCTANE Speed Sandwich, CTC, and CTD products in AWIPS-II. In environments with low-to-marginal vertical wind shear, forecasters were advised to lower the maximum speed value in their color bar to accentuate lower speeds. OCTANE developers also shared an OCTANE Speed Sandwich procedure with a compressed color scale, which was well received by several forecasters for its ability to accentuate speed gradients for monitoring updraft initiation and divergence. A few forecasters also adjusted the CTC and CTD product displays, such as swapping the color tables applied for each product or viewing CTC and CTD separately to improve feature identification. When smoothing the CTD field, the calculated divergence values changed and therefore the color bars for each smoothing field were adjusted so divergence features would appear similar between all three versions. Forecasters were made aware of these changes by the developer, and this frequently led to discussions concerning what cloud-top divergence values for an updraft were operationally significant. These alternative product displays were identified in group discussions, and some found them helpful. However, by increasing the number of display configurations the learning curve for these OCTANE products also increased. Forecasters suggested training materials to communicate product thresholds, and recommended displays through QuickGuides or web-based hands-on activities.

> 'I really like the rainbow compressed color curve [for CTD]. This seemed much easier to interpret, visually, and seemed to work well in both high shear and low shear environments. I suggest keeping this as a color curve option.'
> *NWS Forecaster – End of Week Survey*
>
> 'The OCTANE speed-compressed product [Figure 9] showed a developing updraft with cooling cloud tops and an expanding anvil. There was a tight gradient on the west side of the storm showing good speed divergence. Also, another developing thunderstorm was starting to catch our eye northeast was also growing with good speed divergence…The combination of GREMLIN, OCTANE, CH-13 Clean IR, lightning, LightningCast (not shown) gave us confidence to issue the first severe thunderstorm warning of the day.'
> 4 June 2025, Blog Post: *Day 3 – Using GREMLIN and OCTANE in Warning Operations*
> https://inside.nssl.noaa.gov/ewp/2025/06/04/day-3-using-gremlin-and-octane-in-warning-operations/

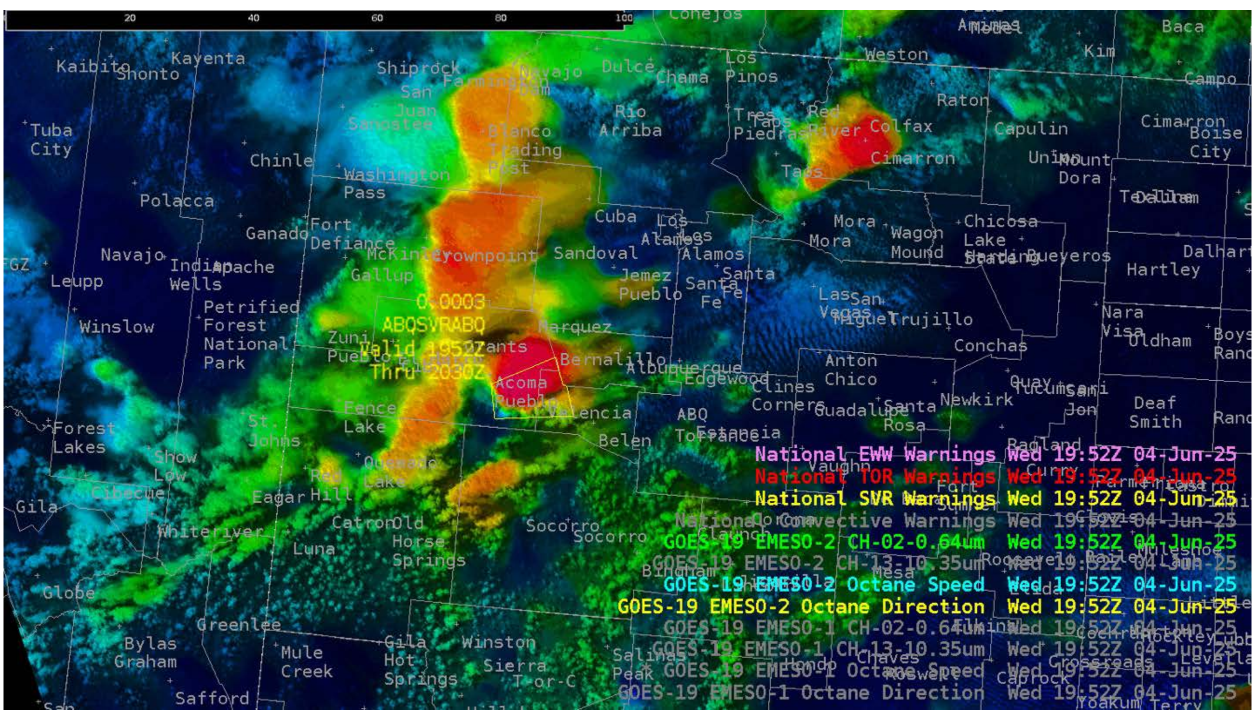

Figure 9: The OCTANE Speed Sandwich product with a compressed color scale of thunderstorms in New Mexico at 1952 Z on 4 June 2025. A severe thunderstorm warning issued at 1952 Z. Animation.

Synthetic imagery from the OCTANE MesoAnywhere product, and their interpretation by forecasters, were compared against ABI Mesoscale scene imagery when available for their CWA. The Day Cloud Phase Distinction RGB from MesoAnywhere was most often used throughout the experiment to monitor developing convection. Forecasters were asked in the daily survey how useful they found the OCTANE MesoAnywhere imagery when compared to the ABI CONUS scene imagery. Approximately 43% (21/49) found MesoAnywhere 'Moderately useful' (level 3 of 5), while 34% responded with 'Very useful' or 'Extremely useful' (levels 4 and 5 of 5) and 22% responded with 'Not at all useful' or 'Slightly useful' (levels 1 and 2 of 5). When observing individual thunderstorms, especially in environments with pronounced vertical wind shear, errors in the MesoAnywhere imagery from ABI Mesoscale scene imagery created wave-like patterns as they readjusted to the ABI CONUS imagery every five minutes. Several forecasters also observed convective features becoming increasingly blurred or distorted, however in group discussions only a few participants found these errors distracting.

Forecasters frequently had access to an ABI Mesoscale scene, however when not available forecasters more frequently used MesoAnywhere for rapidly updating imagery. In group discussions and blog posts forecasters expressed how MesoAnywhere provided a more continuous stream of information. Similarly in the end of week surveys, nearly all (15/18) forecasters felt MesoAnywhere imagery contained little to no or slight differences from ABI Mesoscale scene imagery. Participants were asked to distinguish between various animations of OCTANE MesoAnywhere imagery and ABI Mesoscale scene imagery in the pre-testbed and end of testbed surveys. With exception to one example containing half MesoAnywhere imagery and the other half ABI Mesoscale imagery, most forecasters were able to correctly distinguish between each type of imagery.

> 'MesoAnywhere has proven useful since we do not have a meso sector today. I found that it has been pretty good identifying more dominant storms in decaying clusters with pretty good lead time compared to using 5min imagery. Pretty obvious that this would be useful, but I see it as a pretty great tool as a former Florida WFO meteorologist. A lot can happen in 5 minutes and I see this being quite useful for summertime pulse convection.'
> 21 May 2025, Blog Post: *Using LightningCast and MesoAnywhere for Alligator Wrestling*
> https://inside.nssl.noaa.gov/ewp/2025/05/21/using-lightningcast-and-mesoanywhere-for-alligator-wrestling/

> 'Definitely felt like the MesoAnywhere provided a more high resolution map of the top of the convection whereas the ABI alone felt less detailed. Very small features like overshooting tops were much easier to see with MesoAnywhere than with the ABI data alone. I also liked the 1 minute data, which even if it is synthetic, it gave me a more constant stream of information during warning ops to make my decisions and be continuously thinking aloud to my WFO team.'
> *NWS Forecaster – End of Week Survey*

‘I thought it interesting that while in many instances you don’t see a big change in the MesoAnywhere images from the ‘real’ GOES images at x1 and x6 minutes, vs the interpolated MesoAnywhere imagery, you can tell if you look closely at sharply defined convective clouds with ‘active’ cloud tops full of bubbling and roiling features. The real imagery (example in Fig. 1) will show features that appear somewhat sharper than the interpolated MesoAnywhere imagery [Figure 10]’
22 May 2025, Blog Post: *MesoAnywhere Appearance with Sharply Defined Convective Clouds*
https://inside.nssl.noaa.gov/ewp/2025/05/22/mesoanywhere-appearance-with-sharply-defined-convective-clouds/

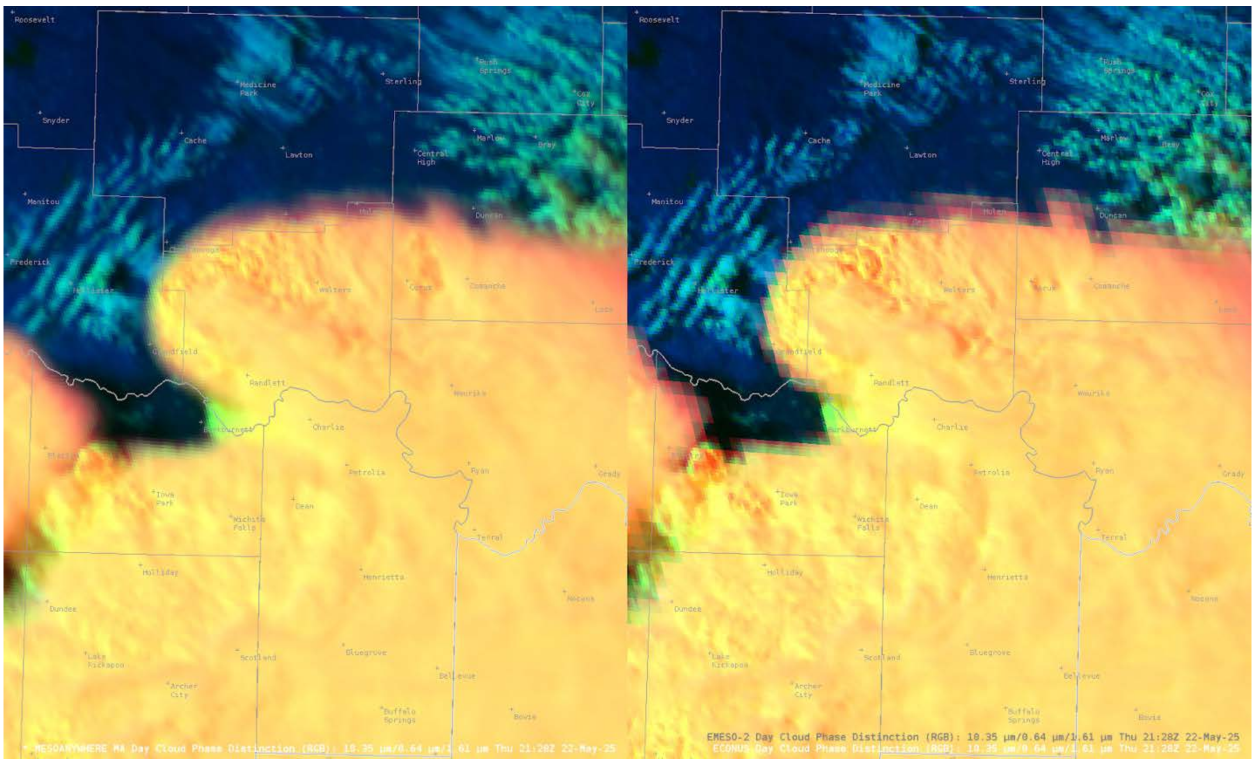

Figure 10: The Day Cloud Phase Distinction RGB of a supercell in southwestern Oklahoma from OCTANE MesoAnywhere (left) and ABI Mesoscale Scene (right) imagery on 22 May 2025.

‘For rapidly developing updrafts, it made the evolution more jumpy in the MesoAnywhere product compared to the ABI Mesoscale data. This didn't so much change my interpretation, but the jumpiness made it less useful. For lower cloud features, though, it seemed to do well, and seemed to offer a nice alternative when ABI mesoscale data isn't available.’
NWS Forecaster – End of Week Survey

Recommendations for Operational Implementation

Based upon the evaluation of the OCTANE suite of products in the 2025 HWT Satellite Proving Ground, the following items have been recommended:

- **It is strongly recommended that low-to-medium levels of smoothing be applied to the OCTANE Cloud-Top Divergence product.** It is expected that divergence signals related to convection will be more clearly depicted due to reduced noise but still retain the ability to depict more localized features related to convective updrafts.
- **It is strongly recommended that NWS forecasters leverage the OCTANE Cloud-Top Cooling and Divergence Sandwich products when interpreting convection initiation and updraft intensity trends.** The ability to quickly identify divergence and cooling signals, especially in environments that marginally support convection, may be considered.
- **It is recommended that alternative display configurations of the OCTANE Speed Sandwich and Cloud-Top Cooling and Divergence Sandwich be explored to optimize OCTANE product applications for a variety of convective scenarios.** The use of compressed or alternative color scales to highlight convective features may be included alongside default options in AWIPS-II and OCTANE-related training materials.
- **It is recommended that NWS forecasters and product developers continue to explore the applications of OCTANE MesoAnywhere imagery, and the ability to provide a more continuous stream of information for diagnosing convective trends, when an ABI Mesoscale scene is unavailable for their area of interest.** Reduced blurring and wave-like effects in scenes with increased vertical wind shear may be explored to increase the utility of these synthetic imagery, along with training to target product applications and limitations.

3.3 Probability of Severe (ProbSevere) LightningCast Model v2

A new version of the NOAA/CIMSS Probability of Severe (ProbSevere) LightningCast model (Cintineo et al. 2022) was evaluated in the 2025 Satellite Proving Ground HWT. The currently operational version of LightningCast (v1) is a machine learning model trained on the reflectance values and brightness temperatures of four predefined ABI bands as input (Band 2: 0.64 µm, Band 5: 1.6µm, Band 13: 10.3µm, and Band 15: 12.3µm) and 60-minute accumulations of the GLM Flash Extent Density (FED) product as truth. Spatial and multi-spectral features from the four GOES-R ABI spectral bands are used in a machine-learning model to predict the probability that the GLM will observe lightning in the next 60 minutes. At night the model input is limited to Band 13 and Band 15. The new experimental version of LightningCast (v2) contains the same ABI inputs and GLM FED target datasets but now includes the MRMS isothermal reflectivity at -10˚C product as input and an expanded training dataset (Cintineo 2025).

In the testbed probabilities from LightningCast v1 and v2 are displayed as contours by default in AWIPS-II with the option for parallax-corrected data, available for the GOES-18/-19, and accessible both day and night. LightningCast v1 and v2 output for static points (stadiums, TAF sites, etc.) are visualized in a time-series format through a web-based lightning dashboard. Forecasters can also leverage lightning dashboards at unique locations and time periods by filling out a request form and generating on-demand dashboards. Forecasters within the testbed were encouraged to request and view LightningCast meteograms in the lightning dashboard tool for each simulated DSS event. With the dashboards, forecasters could assess their effectiveness in identifying and communicating lightning-related information to their partners.

Prior to the testbed, forecasters viewed a training video describing how the ProbSevere LightningCast model generates its output, the definition of the lightning related probabilities, and example applications including forecasting the initiation of lightning and messaging information for decision support. Forecasters received AWIPS-II procedures overlaying the LightningCast v1 and v2 models on the Day Cloud Phase Distinction RGB, GLM FED, and the MRMS Isothermal Reflectivity at -10 °C. The default probability contours were set to 10%, 25%, 50%, and 75% with the ability to modify the contour values and colors by modifying the style rules in the AWIPS-II localization. During the first week of the experiment, LightningCast developers modified these default contours to include the 10%, 30%, 50%, 70%, and 90% probabilities, with a new color table similar the SPC convective outlooks.

ProbSevere LightningCast R2O Questions

- Does the LightningCast model with radar predictors provide more timely and useful information when compared to the model with only satellite predictors? And if so, to what extent?
- From a situational awareness and communication perspective, how important are the lightning dashboards to your IDSS?

Use of ProbSevere LightningCast in the HWT

Of the three products in the 2025 experiment, NWS forecasters were most familiar with LightningCast due to the currently operational version (v1). From the pre-testbed survey, no forecasters marked LightningCast v2 as the product they were most uncertain about, while nearly all (8/10) stated that they used LightningCast frequently in NWS operations. Similar to the currently operational version and its applications in previous testbeds, forecasters consistently reported in the daily survey using LightningCast for decision support (68%, 41/60) and when monitoring for initiating convection and convective maintenance (62%, 37/60). Additionally, half of the forecasters (31/60) used LightningCast to maintain situational awareness for severe convective weather. In the weekly survey all forecasters (16/16) stated they would use LightningCast for decision support, and nearly all (14/16) responded that they would use LightningCast for both situational awareness of initiating convection and for TAF amendments. Lead time for the onset of lightning activity, along with the advection of ongoing activity, from both versions of LightningCast was reported by the forecasters in 10-minute bins from the daily survey (Figure 11). The initiation of lightning activity in the daily survey was spread over multiple bins, with the most responses (11/38) from the '10-19 minutes' bin, followed by the '30-39 minutes' bin (9/38). However, the advection of lightning activity was biased towards the bins with greater lead times, with the '40+ minutes' bin receiving the most responses (13/38) followed by the '30-39 minutes' (11/39) and the '20-29 minutes' (8/39) bins.

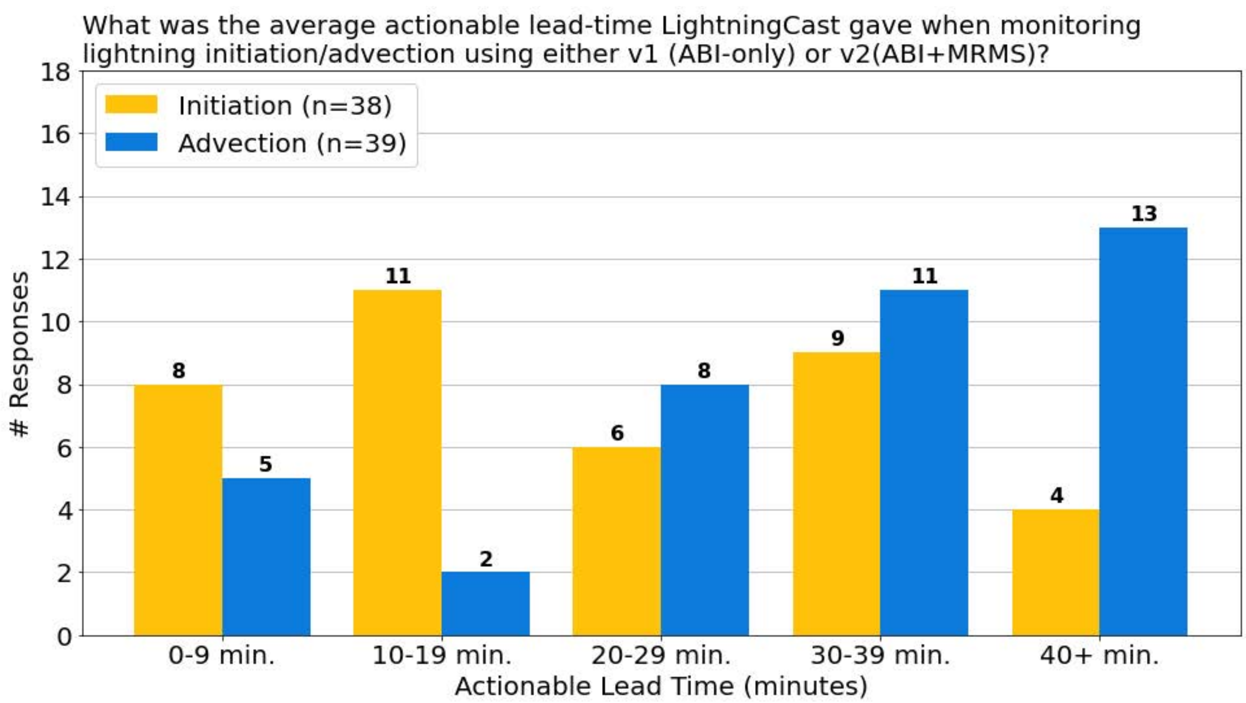

Figure 11: A bar chart showing forecaster reported actionable lead time of ProbSevere LightningCast (v1 or v2) product for the initiation and advection of lightning activity.

Throughout the experiment forecasters used procedures in AWIPS-II and probabilities from the lightning meteogram tool to directly compare the currently operational LightningCast v1 with the experimental LightningCast v2 and observe differences in product accuracy and timeliness. When asked in the daily surveys, over 40% of forecasters (25/59) stated that LightningCast v2 gave longer actionable lead times when compared to LightningCast v1, while vice versa was rarely the reported (3/59). Over one-third of responses (22/59) stated that v1 and v2 were about the same in terms of actionable lead times. In group discussions and blog posts, forecasters remarked several times each week that v2 probabilities would increase – and decrease – sooner and more consistently than v1. This gave the participants increased confidence in thunderstorm development and their ability to communicate lightning-related hazards. Additionally, as was described in the training materials, LightningCast v2 showed notable improvements against LightningCast v1 in scenarios where the satellite signals of convection may be obscured by dense cirrus or adjacent convection. A few participants also mentioned improvements to their timing of lightning cessation for an area or DSS location. Examples of these differences are noted in the following quotes from forecasters.

> ‘V2 decreased in probability faster than V1 on a weakening storm, which helped us key in on the trend and cancel the warning early.’
> *NWS Forecaster – End of Day Survey*

> ‘LightningCast V2 did a great job predicting lightning development with developing convection along a frontal boundary in northwest Iowa. It outperformed version 1, as shown by the loop and images below [Figure 12]…This trend continued throughout, and at 2016Z the first lightning strike was detected. That’s 30 minutes of lead time, which would be helpful for outdoor event IDSS.’
> 2 June 2025, Blog Post: *LightningCast for Convective Initiation and IDSS*
> https://inside.nssl.noaa.gov/ewp/2025/06/02/lightningcast-for-convective-initiation-and-idss/

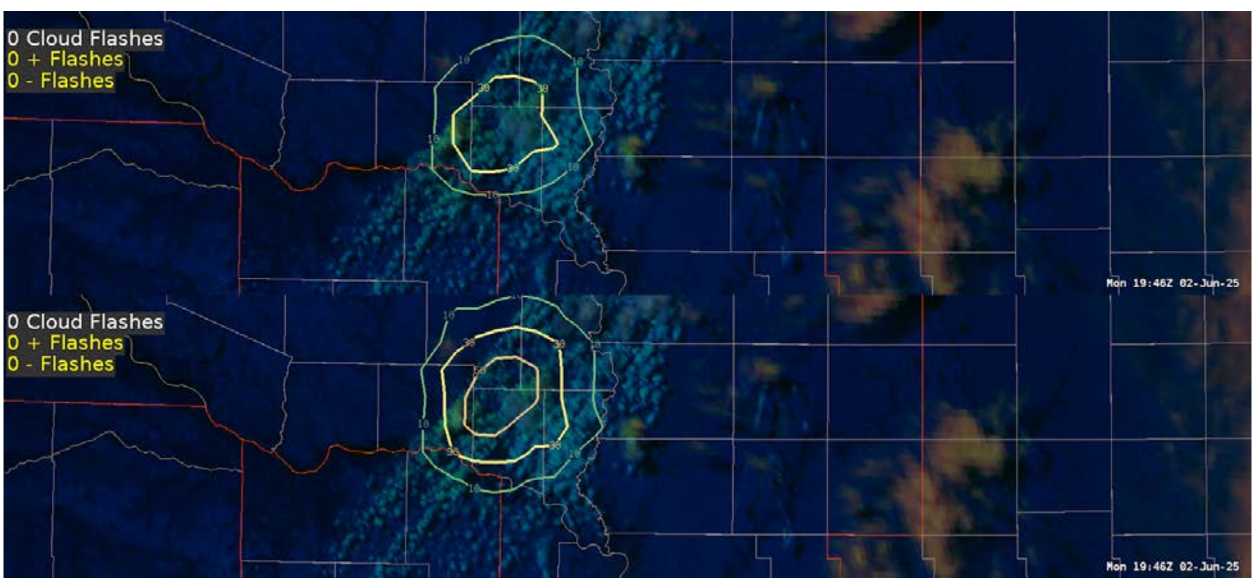

Figure 12: ProbSevere LightningCast v1 (top) and v2 (bottom) overlaid on the Day Cloud Phase Distinction RGB at 1946 Z on 2 June 2025. Animation.

> ‘Though versions 1 and 2 of LightningCast perform generally similarly, I did notice a few times where version 2 captured the initiation of new convection better than version 1,

such as the offshore cells popping up in these screenshots from 19:21 and 19:31Z. [Figure 13]’
21 May, 2025: *LightningCast: Version 1 vs. Version 2*
https://inside.nssl.noaa.gov/ewp/2025/05/21/lightningcast-version-1-vs-version-2/

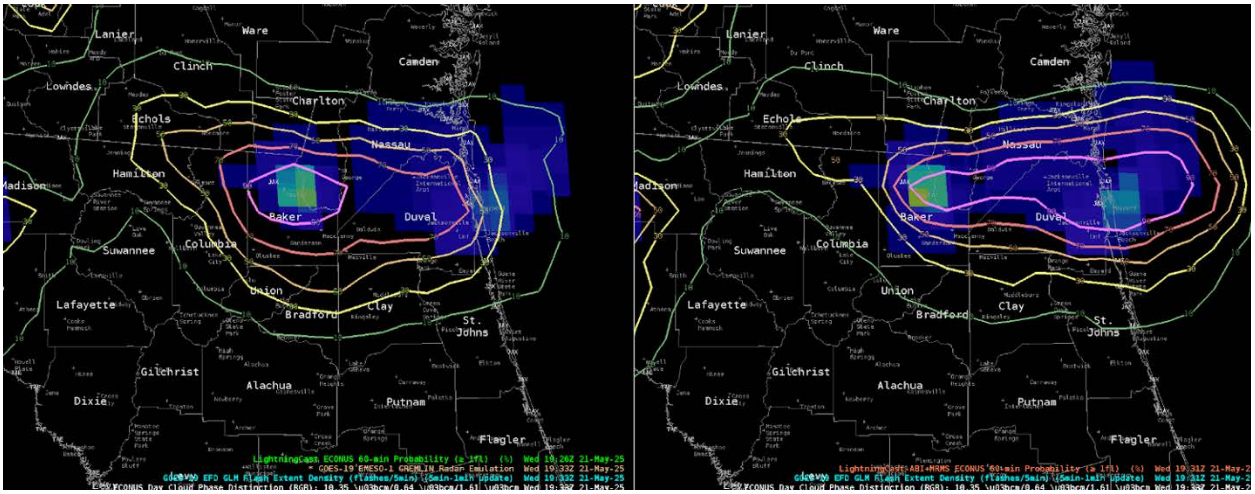

Figure 13: ProbSevere LightningCast v1 (left) and v2 (right) overlaid on GLM Flash Extent Density in northern Florida at 1931 Z on 21 May 2025.

Forecasters stated in group discussions that lightning probabilities were more constrained to the thunderstorms themselves and led to reduced false alarms. Even with this improvement, forecasters also suggested in group discussions the need for more specific timing guidance. The participants supported these comments in the weekly survey when a majority (9/16) stating that a short-fuse lead time (0-30 min.) LightningCast product was ‘Very important/Critical’, the greatest available category. Forecasters during one of the operations days also pointed out locations where higher probability contours from multiple storms would merge, creating probabilities they felt were too high. To further constrain lightning probabilities to the thunderstorm(s), training LightningCast on another ground-based lightning network and a finer horizontal resolution were recommended.

LightningCast v2 frequently showed similar or improved skill in lightning probabilities when compared to LightningCast v1. However, forecasters did observe cases were LightningCast v1 subjectively outperformed LightningCast v2 and these were discussed by the participants. The first case involved rapidly developing convection in an unstable environment with little-to-no inhibition. The impact of including MRMS -10˚C Isothermal Reflectivity in LightningCast v2, and its observational latency of rapid CI compared to satellite signatures of CI, were discussed by the forecasters and developers. In these situations, the probabilities from LightningCast v2 lagged behind LightningCast v1 by a single, 5-minute update. Other cases involved lower LightningCast v2 probabilities in regions with poor radar data. While LightningCast v2 was designed replicate LightningCast v1 in regions outside of the MRMS CONUS domain, its binary classification of radar data quality changed how v2 leveraged this new data source. This effect was most noted along the edge of the MRMS CONUS domain, and in the intermountain-west where radar data coverage around the -10˚C isotherm was limited. In the group discussions, it was suggested that

LightningCast v2 could include training information such as the MRMS Radar Quality Index. The following quotes from forecasters highlight the aforementioned scenarios.

> ‘it appears that v1 does better in areas with poor radar coverage, while v2 does better in areas with better radar coverage. In the image above, version 1 has a better handle on the isolated first GLM pixel (50%) than version 2 (10%). Meanwhile, the more robust lightning area is more accurately represented on version 2 (which happens to have better radar coverage) compared to version 1.’ [Figure 14]
> 2 June 2025, Blog Post: *DSS LightningCast Dashboard*
> https://inside.nssl.noaa.gov/ewp/2025/06/02/dss-lightningcast-dashboard/

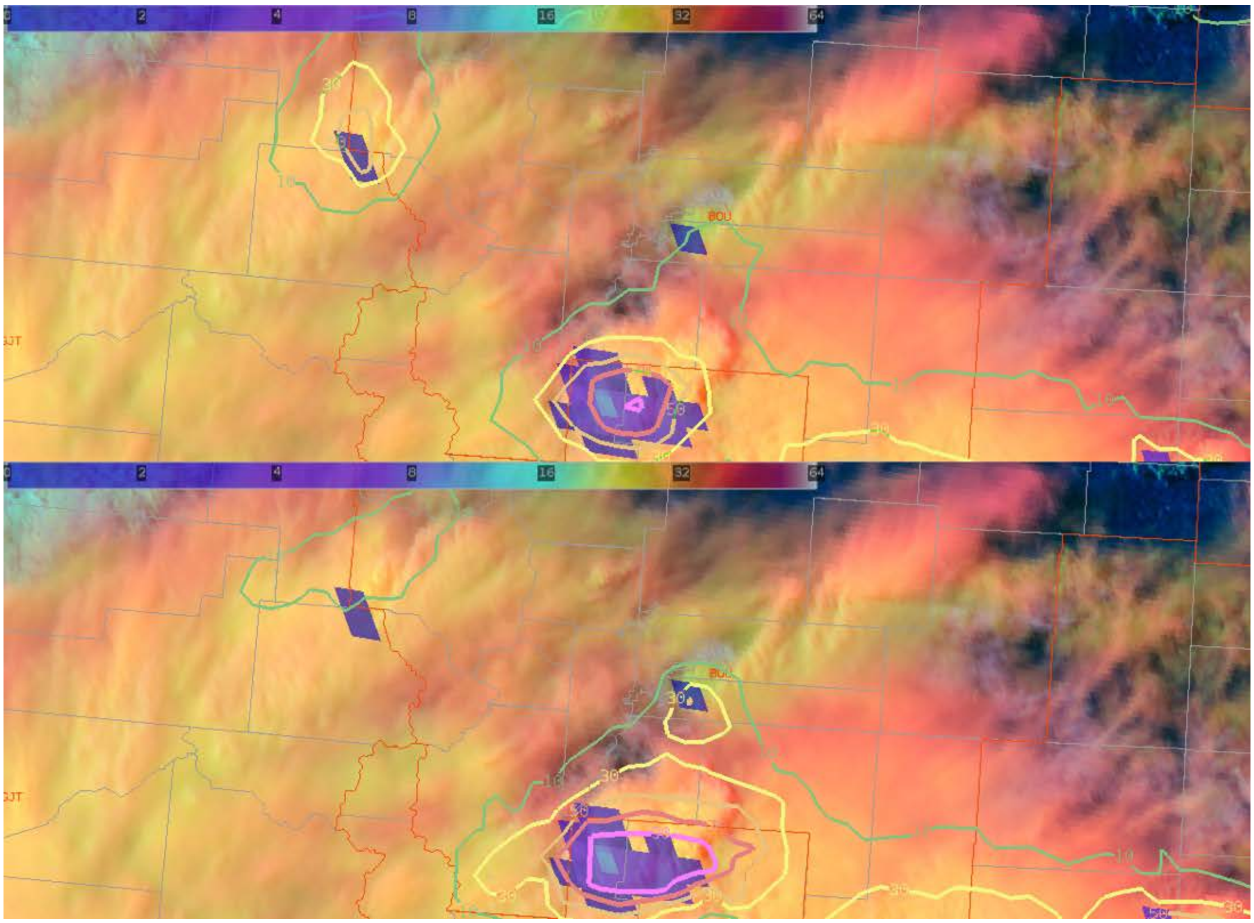

Figure 14: LightningCast v1 (top) and LightningCast v2 (bottom) overlaid on GLM Flash Extent Density and the Day Cloud Phase Distinction RGB in central Colorado at 2221 Z on 2 June 2025. Note the differences in v1 and v2 in the northwest region with poor radar coverage and the central region with good radar coverage in the display.

> ‘Storms today again seemed to be very efficient and quick at producing lightning as they grew vertically, so in many cases (but not all), V1 seemed to be one scan faster than V2 on increasing lightning probabilities for new convection.’
> *NWS Forecaster – End of Day Survey*

‘It seemed like many of the cells were producing lightning very quickly after achieving even modest returns at the -10C level. In this case, the significant instability in the region today led to very rapid vertical cloud growth, which likely outpaced significant radar returns aloft…The V1 product, being based entirely on satellite data, was able to key in on just the rapid vertical growth and boost probabilities based on that alone, and not have to wait for the -10C returns to show up in MRMS.’ [Figure 15]
19 May 2025, Blog Post: *LightningCast Comparison*
https://inside.nssl.noaa.gov/ewp/2025/05/19/lightningcast-comparison/

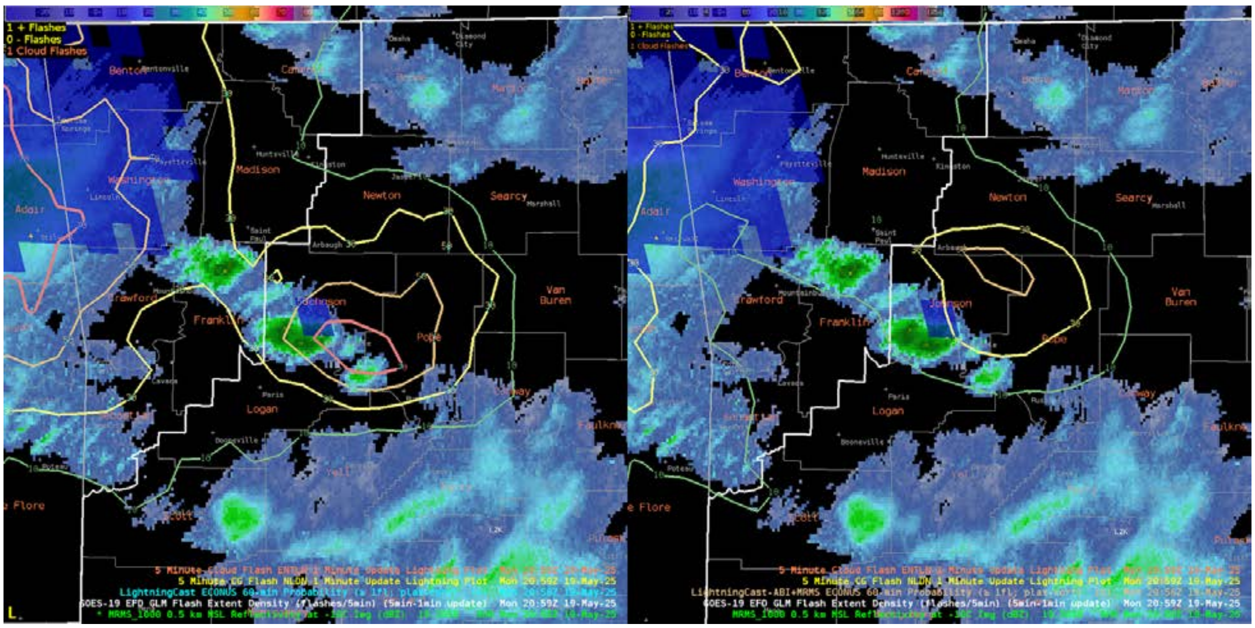

Figure 15: ProbSevere LightningCast v1 (left) and v2 (right) overlaid on GLM Flash Extent Density and MRMS -10˚C Isothermal Reflectivity in Arkansas at 2059 Z on 19 May 2025. Animation.

Along with observing LightningCast in AWIPS-II, forecasters leveraged static and on-demand lightning dashboards to communicate lightning-related hazards in simulated DSS events. Of the 52 messages sent during these simulated DSS events, over two-thirds (37/52) mentioned LigthningCast as an influence in their messaging and 11 mentioned the lightning dashboard specifically. Along with comparing LightningCast v1 and v2, forecasters also discussed with developers their ability to directly communicate probabilities and trends for their DSS events. When asked in the weekly survey how often they would use the static or on-demand dashboards in operatically applicable scenarios, over two thirds responded with the two highest categories of ‘Always or almost always’ or ‘Regularly’.

‘The decision to contact the partner about the DSS event at 1955 UTC was made with the help of the LightningCast DSS Dashboard, which had a max probability of lightning within the 10 mile radius of the event at 90% at the time of the contact. They were told they had less than 30 minutes before lightning was within 10 miles, and 20 minutes after that call, the first GLM strike was observed in that radius.’ [Figure 16]
5 June 2025, Blog Post: *Southwest Texas Lightning Product Performance*

https://inside.nssl.noaa.gov/ewp/2025/06/05/southwest-texas-lightning-product-performance/

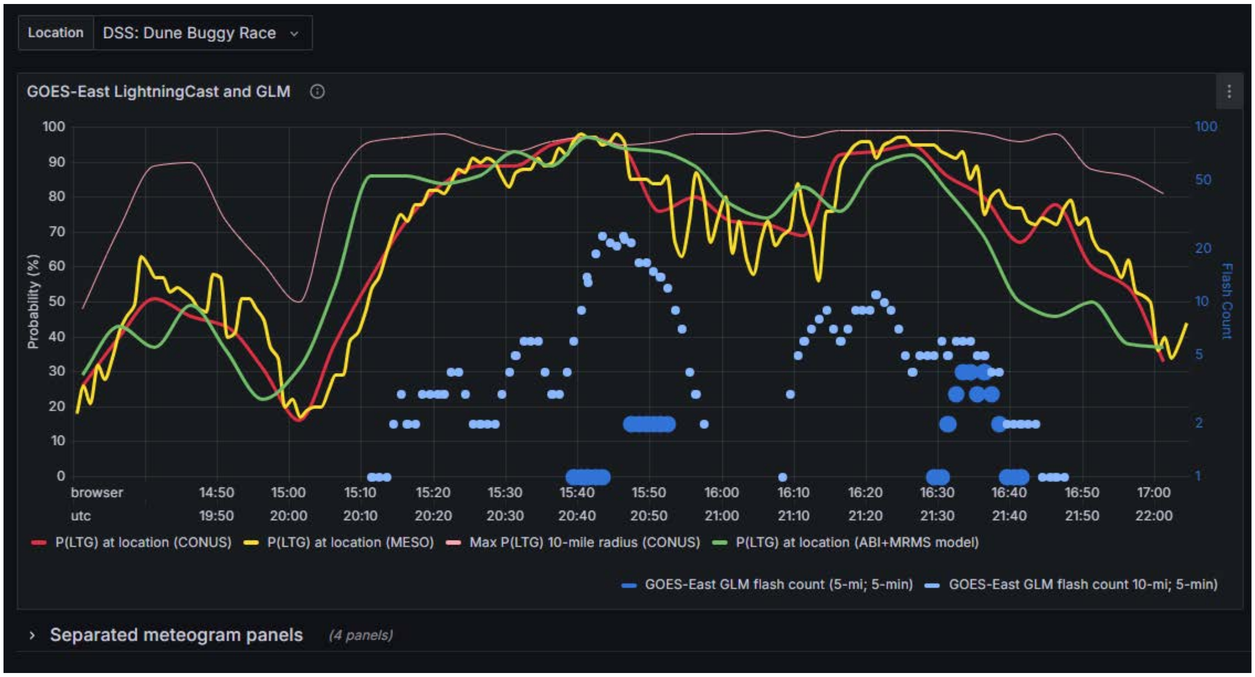

Figure 16: LightningCast probabilities and GLM flash counts from a simulated DSS event on 5 June 2025.

While well received, forecasters consistently made suggestions to improve the lightning dashboards. The most common suggestion from forecasters was to customize the 10-mile maximum probability in the on-demand dashboards based on a user-specified range. A 10-mile range was considered a good starting point for most events, however the forecasters said this range can be as short as 5 miles or as long as 20 miles based on partner needs. Forecasters also suggested the ability to put multiple on-demand dashboards in a single, multi-panel display to improve situational awareness when more than one event is ongoing. Within the dashboards, forecasters learned how to toggle individual time series from the meteograms, isolate individual panels, and use the available keyboard shortcuts. A few forecasters expressed the difficulty of finding these functions, and suggested training in the form of QuickGuides or training videos may increase the accessibility of the dashboard for new users. Lastly, inset maps of LightningCast probabilities around the user-defined site were mentioned as part of the dashboard to provide both temporal and spatial context for lightning probabilities.

Another new feature of the LightningCast display in AWIPS-II for the experiment was a change in the number and color of the probability contours. The five contour colors were selected and ordered to match those used by the SPC for convective outlooks. Forecasters readily accepted this modified display configuration, and most expressed that it made communicating product displays easier as it matched many of their current color scales for intensity. This was reflected in the consistent use of LightningCast probabilities by graphics created for the simulated DSS events (Figure 17). In one case, forecasters created a version of LightningCast with filled contours which could then be used in their graphical messages (Figure 18).

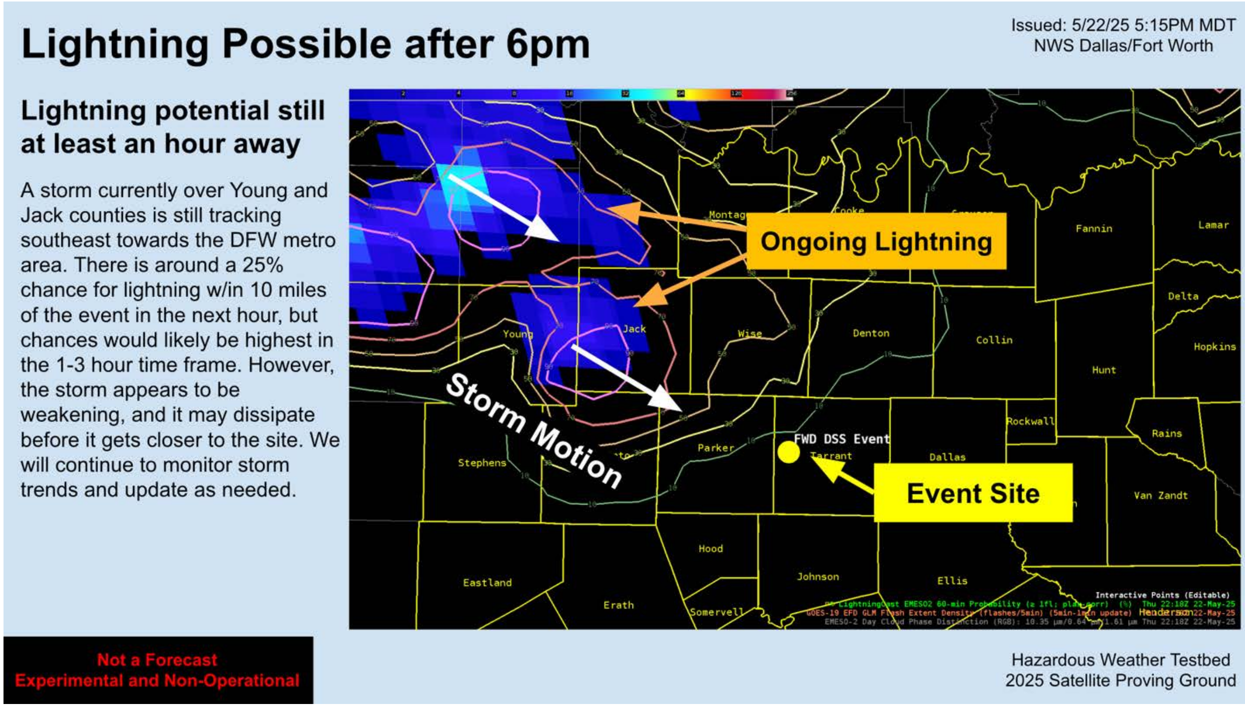

Figure 17: Forecast graphic for a simulated DSS event on 22 May 2025 in Northern Texas, featuring LightningCast v1 probability contours and GLM Flash Extent Density.

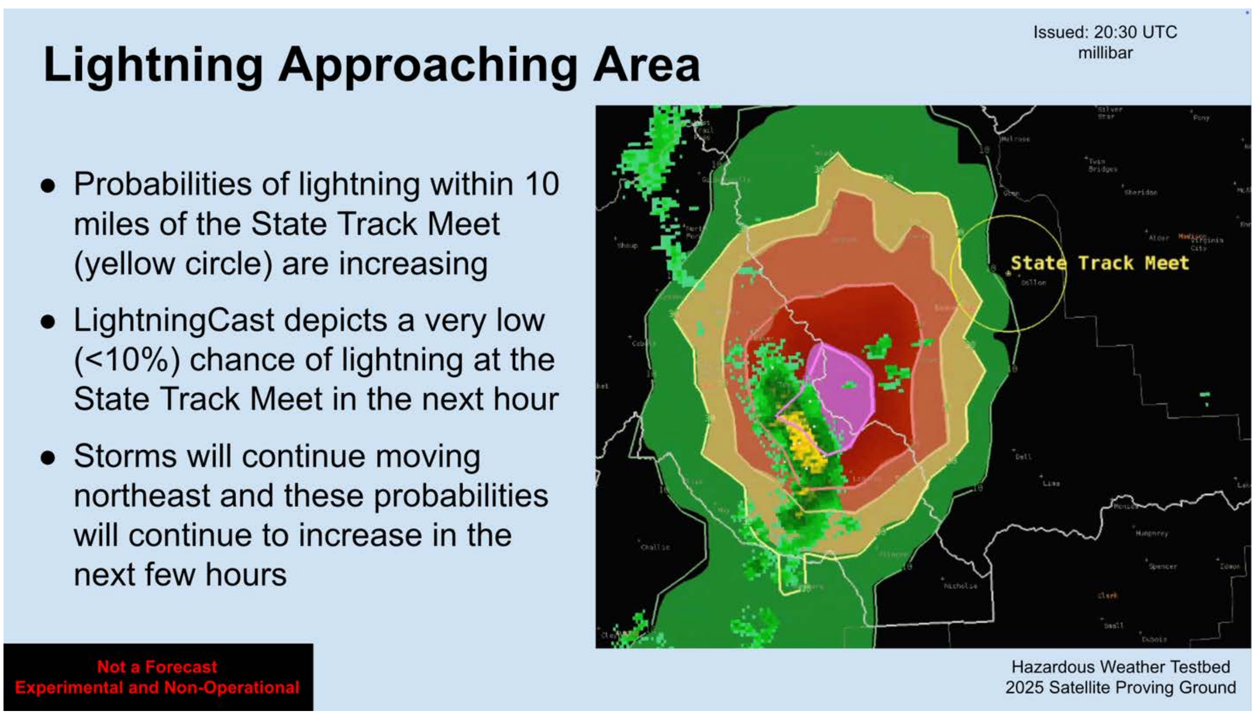

Figure 18: Forecast graphic for a simulated DSS event on 22 May 2025 in western Montana, featuring LightningCast v1 probability contours and radar reflectivity.

Recommendations for Operational Implementation

Based upon the evaluation of ProbSevere LightningCast in the 2025 HWT Satellite Proving Ground, the following items have been recommended:

- **It is strongly recommended that NWS forecasters leverage the ProbSevere LightningCast Version 2 model to provide overall more timely and accurate probabilistic lightning information, especially for the development and cessation of lightning activity.** The included radar information provides more reliable and timely probabilities in a vast majority of scenarios where LightningCast Version 1 probabilities were limited in their utility, based on previous demonstrations in the HWT.
- **It is recommended that the ProbSevere LightningCast Version 2 model explore further improvements to its probabilistic lightning information to increase its operational impact and utility.** Suggested avenues of development and improvement include new model predictors for interpreting MRMS data quality, improved timing guidance, varied or multiple lightning data sources for model training, and increased sensitivity for rapidly developing convection.
- **It is strongly recommended that the ProbSevere LightningCast lightning dashboards are applied by NWS forecasters to efficiently communicate lightning probabilities for impact-based decision support, situational awareness, and public safety messaging.** Expanded functionality such as user-specified range-rings and viewing multiple dashboard points on a single webpage are suggested. Training to prepare new users with the dashboard interface was also suggested in the form of a one-page document (e.g. QuickGuides) or a training video.
- **It is strongly recommended that ProbSevere LightningCast contours leverage a color table similar to the Storm Prediction Center convective outlook, to provide consistent messaging with other NWS-related color tables.**

4. Summary and Conclusions

The GOES-R Satellite Proving Ground conducted three weeks of remote product evaluations during the 2025 Spring Experiment of the Hazardous Weather Testbed. Seventeen NWS forecasters evaluated three experimental GOES-R products and interacted with their algorithm developers and visiting scientists during the experiment. Quantitative feedback was collected through surveys administered at the end of each day and week, along with a warning/decision support service reporting form. Qualitative feedback came from daily discussions with forecasters, blog posts, and public graphics. Along with the standard warning operations, mock decision support service events were created most days to reflect the full range of intended operational applications. Participants received training materials prior to the testbed for each product through a combination of user guides, PowerPoint presentations, and training videos. Products were also summarized at the beginning of each week by their developers, which included product applications, limitations, locations in AWIPS-II, and recommended display practices.

This year's experiment featured two new aspects, including radar data-denial events and a pre-testbed survey. In the radar-denial events, groups were either placed in regions with poor radar coverage or instructed to simulate a radar outage for their location by not viewing radar data for the entire demonstration case. These events were intended to challenge the forecasters to fully exploit these experimental tools for convective warnings and decision support. Pre-testbed surveys were sent to the forecasters before their week in the experiment, as an optional activity. Forecasters provided advanced insight into the products they were about to demonstrate, on topics such as their familiarity with currently operational products, their current techniques for handling scenarios that the experimental products were expected to address, and their initial confidence in the experimental products based on the training materials.

The 2025 satellite proving ground experiment completed 12 days – or approximately 56 hours – of simulated, live product demonstrations between 5 May and 6 June 2025. Forecaster groups were localized to 22 unique Weather Forecast Offices across the continental United States following the daily regions of expected hazardous convection (Figure 19). Forecasters issued 146 convective products using WarnGen, with 121 Severe Thunderstorm Warnings, 16 Tornado Warnings, and 9 Special Weather Statements (Figure 20). From the 22 mock decision support events created in the experiment, forecasters sent 52 messages and 27 graphics to their partners to communicate hazard probability, timing, and intensity. In providing feedback related to their experiences in the testbed, forecasters also completed 68 blog posts, 10 pre-testbed surveys, 61 daily surveys, 17 weekly surveys, and 15 hours of group discussions with product developers and visiting scientists. All of the collected data from this experiment provided insight into how forecasters applied GREMLIN, OCTANE, and ProbSevere LightningCast v2 along with their feedback and recommendations.

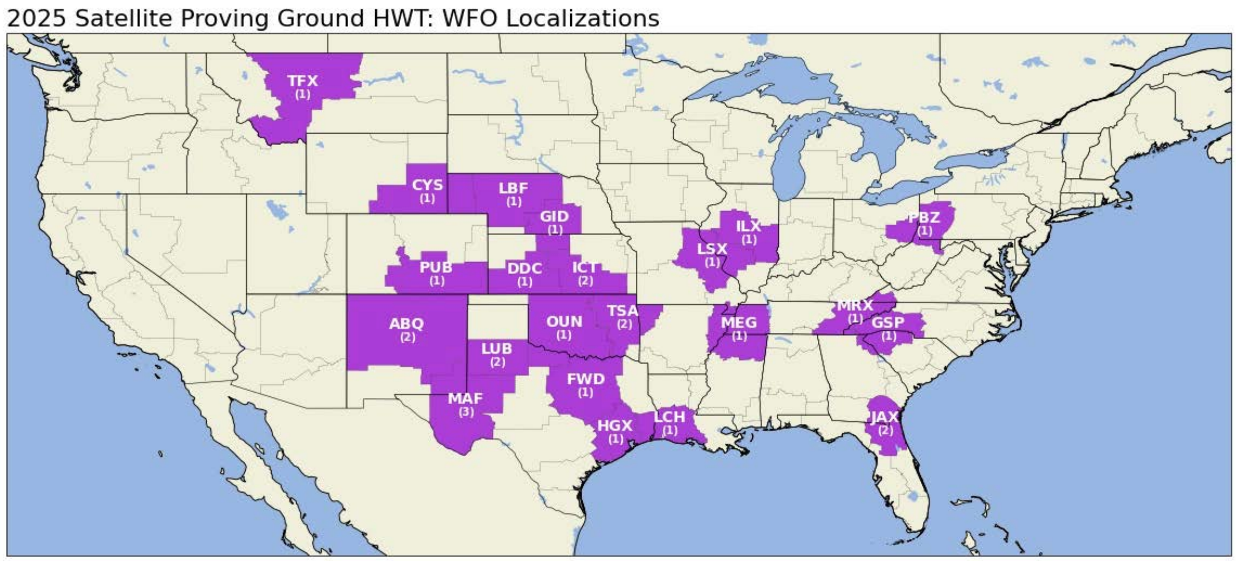

Figure 19: A map of NWS WFOs the 2025 experiment were localized to (purple), and the number of times forecasters were localized to each office.

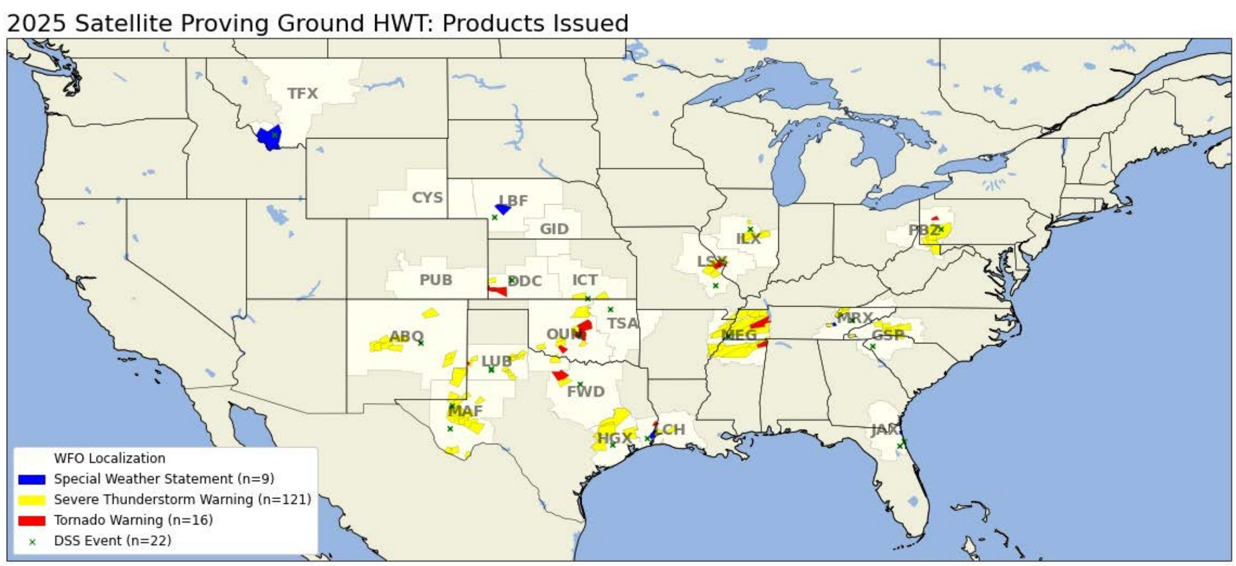

Figure 20: A map of all convective products issued by forecasters in the 2025 experiment, including Special Weather Statements (blue), Severe Thunderstorm Warnings (yellow), and Tornado Warnings (red). Mock decision support service events are noted in green.

Along with feedback from the experimental products themselves, forecasters also shared their opinions on the experiment and recommendations for potential improvement. Overall forecasters enjoyed their time in the testbed, when asked in the end-of-week discussion and weekly survey. The ability for NWS forecasters, satellite product developers, and visiting scientists to interact in a collaborative learning environment was praised by all groups in attendance. A few forecasters also mentioned the ability to learn from the participants from other offices and regions regarding convective warning operations and decision support. Developers and visiting scientists expressed that they departed the experiment with direct feedback from their users and potential avenues for future research studies and product development initiatives. This feedback highlighted the value of annual Satellite Proving Ground activities in the Hazardous Weather Testbed for the research and operational communities.

> 'The Satellite HWT is a great way to assist R2O by providing feedback on potential new satellite products. You have the unique opportunity to have one on one discussions with the developers while testing the products in real-time operations.'
> *NWS Forecaster – End of Week Survey*

> 'My time in the satellite HWT was fun and worthwhile. It was nice to have dedicated time to test out products and more importantly, discuss the products with developers and facilitators at HWT. I enjoyed working with forecasters in different offices and hearing how they would go about a product, DSS post, mesoanalysis, etc.'
> *NWS Forecaster – End of Week Survey*

> 'I really enjoyed the testbed. I was impressed to see updates to all of the products evaluated last year. Mesoanywhere, updates to LightningCast, and the new 5-minute GREMLIN product are all products I think would be relevant to operations that I'll take back.'
> *Visiting Scientist – End of Testbed Survey*

Along with product assessments, forecasters also shared feedback on the organization of the experiment and provided several recommendations for its improvement for future iterations. During the first week of the experiment all participants recommended additional guidance for daily testbed logistics beyond what was provided in the demonstration plan, such as a one-pager document or 'crash course' training video. A one-page document was created before the second week of the experiment to outline the experimental satellite products, the functionality of AWIPS-II, and key logistics within operations. Forecasters from the second and third week found this document helpful when preparing for their week in the testbed. The real-time nature of the experiment was well received. Forecasters felt this setup replicated operations and the variety of convective scenarios they may encounter. On 'slow' days with less intense convection, forecasters and developers suggested an archived case may still be useful. Additional topics raised included the time spent creating and modifying procedures at the start of the week, fatigue on days with copious convection across their CWA, more efficient ways to share information outside of blog posts, and the limited number of templates available when creating graphics for simulated DSS events.

> 'I think the layout of the experiment was great and really liked that we were allowed to take breaks as needed. I would recommend keeping the severe RAC procedures in the AWIPS cloud for future participants so they do not have to wait time making procedures. A few things were missing from the procedures which would have been helpful to have for warning operations.'
> *NWS Forecaster – End of Week Survey*

While beneficial, forecasters and product developers alike noted difficulties when collaborating and communicating in an all-virtual experiment. Forecasters at times faced network issues that limited the functionality of AWIPS-II in their cloud-based instances, were limited in screen space to view products, couldn't interact with the forecasters in other groups during the demonstration periods, and faced fatigue multi-hour video calls when in their breakout groups. Product developers also expressed difficulty in monitoring up to three forecaster breakout groups during operations, along with sharing information with forecasters as questions and discussion topics arose. When proposing potential configurations of future experiments to product developers, the option with all three weeks held virtually was the least popular with all responding with 'Moderately oppose' or 'Strongly oppose'. The 2024 experiment format (2 in-person weeks, 1 virtual week) was the most supported, with none of the respondents in opposition. These results are similar to developer feedback collected in the 2024 experiment.

> 'The most challenging aspect of the experiment is the virtual nature, which I know can't easily be improved. It's challenging in this environment to realistically simulate WFO operations, especially with the limited screen real estate.'
> *NWS Forecaster – End of Week Survey*

> 'In-person experiment weeks provide far more valuable feedback and foster more lasting relationships between end-users and science developers and should be the preferred format whenever possible.'
> *Product Developer – End of Testbed Survey*

> 'This year's all in-person HWT was still a valuable experience that provided useful information for R2O and O2R. But my best experience was last year with in-person... being able to look over a forecasters shoulder, being able to talk over dinner - this provides information and allows for making connections that are really difficult in the virtual format. At times it could be really difficult to monitor and engage with three breakout rooms.'
> *Product Developer – End of Testbed Survey*

Recommendations for Future GOES-R HWT Satellite Proving Ground Experiments

Based upon experiences and feedback from the 2025 GOES-R HWT Satellite Proving Ground, the following recommendations for future testbeds are included below:

- It is highly recommended future Satellite Proving Ground HWT experiments feature in-person demonstration portions to fully exploit the collaborative nature of the experiment between the NWS operational and research communities. A format similar to the 2024 experiment (2 in-person weeks, 1 virtual week) may be strongly considered.
- It is recommended that the Satellite Proving Ground HWT modify the experiment design, such as splitting each day during the virtual week(s) into two sessions, flexible start times for operations, archived cases, and other methods for forecasters to submit their feedback aside from blog-style posts.

4.1 Acknowledgements

Multiple NWS meteorologists participated in this experiment and provided detailed feedback that went into the recommendations from this report. It would not be possible to complete to research-to-operations and operations-to-research cycle without their willingness to work with experimental data, ask questions, and participate in crucial discussions with product developers. The developer teams themselves, along with visiting scientists, have continued to leverage the experiment as a collaborative environment to test emerging technologies and transfer impactful knowledge to NWS operations. Justin Monroe and Johnathan Madden (OU CIWRO) provided exceptional technical support during all phases of the experiment, which allowed forecasters to effectively use the cloud-based AWIPS-II instances for the evaluation. Anthony Lyza (NSSL) provided logistical oversight for the Hazardous Weather Testbed. The persons listed below served as product developers. The collaborative efforts of NWS forecasters, product developers, and visiting scientists all play a crucial role in the impact of the experiment. The SPC and HWT Satellite Liaison thanks them for their contributions.

GREMLIN

Kyle Hilburn	Colorado State University – CIRA

OCTANE

Jason Apke	Colorado State University – CIRA
Bill Line	NOAA – NESDIS

ProbSevere LightningCast

Lena Heuscher-Stewart	University of Wisconsin – CIMSS
John Cintineo	NOAA – NSSL
Justin Sieglaff	University of Wisconsin – CIMSS
Michael Pavolonis	NOAA – NESDIS

5. Appendix

5.1 List of Abbreviations

ABI	Advanced Baseline Imager
AWIPS-II	Advanced Weather Information Processing System – Version 2
CIMSS	Cooperative Institute for Meteorological Satellite Studies
CIRA	Cooperative Institute for Research in the Atmosphere
CISESS	Cooperative Institute for Satellite Earth System Studies
CIWRO	Cooperative Institute for Severe and High-Impact Weather Research and Operations
CONUS	Continental United States
CTC	Cloud-Top Cooling
CTD	Cloud-Top Divergence
CWA	County Warning Area
DCP	Day Cloud Phase Distinction
DSS	Decision Support Service
EWP	Experimental Warning Program
GLM	Geostationary Lightning Mapper
GOES-R	Geostationary Operational Environmental Satellites R-Series
GREMLIN	GOES Radar Estimation via Machine Learning to Inform NWP
HWT	Hazardous Weather Testbed
JPSS	Joint Polar Satellite System
NOAA	National Oceanic and Atmospheric Administration
NSSL	National Severe Storms Laboratory
NWS	National Weather Service
MESH	Maximum Expected Size of Hail
MRMS	Multi-Radar/Muli-Sensor
NESDIS	National Environment, Satellite, Data, and Information Service
OCTANE	Optical flow Code for Tracking, Atmospheric motion vector, and Nowcasting Experiments
OU	University of Oklahoma
O2R	Operations-to-Research
RGB	Red-Green-Blue
R2O	Research-to-Operations
SPC	Storm Prediction Center
TAF	Terminal Aerodrome Forecast
VII	Vertically Integrated Ice
VIL	Vertically Integrated Liquid
WFO	Weather Forecast Office
WSR-88D	Weather Service Radar – 1988 Doppler

6. References

Apke, J. M., Y.-J. Noh, and K. Bedka, 2022: Comparison of Optical Flow Derivation Techniques for Retrieving Tropospheric Winds from Satellite Image Sequences. *Journal of Atmospheric and Oceanic Technology*, **39**, 2005–2021, https://doi.org/10.1175/JTECH-D-22-0057.1.

Calhoun, K. M., K. L. Berry, D. M. Kingfield, T. Meyer, M. J. Krocak, T. M. Smith, G. Stumpf, and A. Gerard, 2021: The Experimental Warning Program of NOAA's Hazardous Weather Testbed. *Bulletin of the American Meteorological Society*, 1–51, https://doi.org/10.1175/BAMS-D-21-0017.1.

Cintineo, J. L., 2025: The Impact of MRMS in a Satellite-Based Lightning Nowcasting Model.

——, M. J. Pavolonis, and J. M. Sieglaff, 2022: ProbSevere LightningCast: A Deep-Learning Model for Satellite-Based Lightning Nowcasting. *Weather and Forecasting*, **37**, 1239–1257, https://doi.org/10.1175/WAF-D-22-0019.1.

Hilburn, K. A., I. Ebert-Uphoff, and S. D. Miller, 2021: Development and Interpretation of a Neural-Network-Based Synthetic Radar Reflectivity Estimator Using GOES-R Satellite Observations. *Journal of Applied Meteorology and Climatology*, **60**, 3–21, https://doi.org/10.1175/JAMC-D-20-0084.1.

Thiel, K., 2023: *GOES-R and JPSS Proving Ground Demonstration at the Hazardous Weather Testbed 2023 Spring Experiment Final Evaluation*. Cooperative Institute for Severe and High-Impact Weather Research and Operations, accessed 15 November 2023, https://doi.org/10.25923/37va-6g77.

——, 2024: *GOES-R and JPSS Proving Ground Demonstration at the Hazardous Weather Testbed 2024 Spring Experiment*. Cooperative Institute for Severe and High-Impact Weather Research and Operations, https://doi.org/10.25923/xze9-r006.

This report was prepared by Kevin Thiel with funding provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA21OAR4320204, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of NOAA or the U.S. Department of Commerce.